DEVELOPMENT OF QUANTUM-CRYSTALLOGRAPHIC METHODS FOR CHEMICAL AND BIOCHEMICAL APPLICATIONS

Inaugural dissertation
of the Faculty of Science,
University of Bern

presented by

FLORIAN KLEEMISS

from **Bremen, Germany**

Supervisor of the doctoral thesis:

PD Dr. Simon Grabowsky

Department of Chemistry and Biochemistry
University of Bern

DEVELOPMENT OF QUANTUM-CRYSTALLOGRAPHIC METHODS FOR CHEMICAL AND BIOCHEMICAL APPLICATIONS

Inaugural dissertation
of the Faculty of Science,
University of Bern

presented by

FLORIAN KLEEMISS

from **Bremen, Germany**

Supervisor of the doctoral thesis:

PD Dr. Simon Grabowsky

Department of Chemistry and Biochemistry
University of Bern

Accepted by the Faculty of Science.

Bern, 14.12.2020

The Dean
Prof. Dr. Zoltán Balogh

To the most important person, my wonderful and loving partner,
in perpetuum et unum diem.

Contents

List of Abbreviations v

List of Figures vii

List of Tables x

Abstract xiii

1 Introduction 1

1.1 Brief History and Introduction to Crystallography 1
  1.1.1 Crystallography 1
  1.1.2 Diffraction Experiment 1
  1.1.3 Radiation sources 2
  1.1.4 Detectors 3
  1.1.5 Structure Solution and Refinement 4
  1.1.6 X-ray Data Quality and Resolution 4
  1.1.7 Independent Atom Model (IAM) 6
  1.1.8 Non-Spherical Refinements - History and Models 8
  1.1.9 Wavefunction Based Refinements 10
1.2 Quantum Crystallography - A Brief Definition 10
  1.2.1 Hirshfeld Atom Refinement, X-ray Constrained Wavefunction Fitting and X-ray Wavefunction Refinement 12
  1.2.2 Hydrogen Atoms in XWR 13
  1.2.3 Restrictions in Wavefunction Based Refinements 14
1.3 Complementary Bonding Analysis 16
  1.3.1 Real Space Descriptors 16
  1.3.2 Orbital Space Analysis 18
  1.3.3 Energy Space Analysis 18
1.4 Applying Quantum Crystallography in Drug Design 18
  1.4.1 Protein-Ligand Interaction 18
    1.4.1.1 Key-Lock Principle 19
    1.4.1.2 Induced-fit Mechanism 19
    1.4.1.3 Modern Protein-Ligand Analysis 20
  1.4.2 Rational Drug Design 20
    1.4.2.1 Using Theoretical Models 20
    1.4.2.2 Using Crystallography 22
1.5 Cysteine Proteases 24
1.6 Ibuprofen 24
1.7 Carbon-Silicon Switch 25
1.8 Examples of Protein-Ligand Interactions in this Thesis 26

2 **Theoretical Background** **29**
2.1 Crystallography . . . 29
2.1.1 Measurement . . . 29
2.1.2 Diffractometer . . . 30
2.1.3 Evaluation of Experimental Data . . . 31
2.1.4 Structure Solution . . . 33
2.1.5 Least-squares Refinement . . . 36
2.1.6 Independent Atom Model (IAM) . . . 37
2.1.7 Multipole Model . . . 37
2.1.8 Hirshfeld Atom Refinement (HAR) . . . 37
2.1.9 X-ray Constrained Wavefunction Fitting (XCW) . . . 38
2.1.10 X-ray Wavefunction Refinement (XWR) . . . 40
2.2 Wavefunction Calculations . . . 40
2.2.1 Approximations . . . 41
2.2.2 Basis Sets . . . 43
2.2.3 Methods . . . 44
2.2.3.1 Self-Consistent Field (SCF) . . . 44
2.2.3.2 Hartree-Fock (HF) . . . 44
2.2.3.3 Electron Correlation (EC) . . . 45
2.2.3.4 Density Functional Theory (DFT) . . . 46
2.2.3.5 Density Fitting / Resolution of Identity (RI) . . . 49
2.2.3.6 Chain of Spheres (COSX) . . . 50
2.3 Bonding and Interaction Indicators . . . 50
2.3.1 Quantum Theory of Atoms in Molecules (QTAIM) . . . 51
2.3.2 Interaction Density . . . 53
2.3.3 Non-Covalent Interactions (NCI) . . . 53
2.3.4 Electrostatic Potential (ESP) . . . 55
2.3.5 Electron Localizability Indicator (ELI) . . . 57
2.3.6 Interaction Energies . . . 58
2.3.7 Natural Bond Orbitals (NBO) . . . 59
2.3.8 Roby-Gould Bond Index (RGBI) . . . 59
2.3.9 Calculation of Numerical Data in Parallel . . . 60
2.4 Molecular Dynamics (MD) . . . 60
2.4.1 Classical Force Field . . . 61
2.4.2 Free-Energy-Perturbation Methods (FEP) . . . 65
2.4.3 Root Mean Square Displacement (RMSD) . . . 66
2.5 Quantum Mechanics Molecular Mechanics Hybrid Methods (QM/MM) . . . 66

3 **Enabling Quantum Crystallography** **69**
3.1 Determination of Accurate Atom Parameters from X-ray Diffraction using Non-Spherical Atomic Form Factors . . . 69
3.1.1 Using HAR in *Tonto* . . . 69
3.1.2 Implementation in *Olex2* as *NoSpherA2* . . . 70
3.1.2.1 The *.tsc* File . . . 70
3.1.2.2 Principle of HAR in NoSpherA2 . . . 71
3.1.2.3 Validation of NoSpherA2 . . . 72
3.1.2.4 Relativistic Treatment . . . 73
3.1.2.5 Disorder . . . 73
3.1.2.6 Choosing a Level of Theory for HAR . . . 76
3.1.3 *Publication:* Accurate Crystal Structures and Chemical Properties from NoSpherA2 . 78

3.2 Averaged Interaction Energies (aIE) 97
3.3 The CUDA-accelerated Quantum-Crystallographic-Toolbox (cuQCrT) 99
3.3.1 Preparation and Conversion of Wavefunction Files 99
3.3.2 Real Space Properties 100
3.3.3 Bond Based Calculation of Properties 102
3.3.4 Orbital Space Analysis 103
3.3.5 X-ray Constrained Complementary Bonding Analysis 103

4 Applying Quantum Crystallography for Complementary Bonding Analysis 107
4.1 Experimental Electron Correlation, Exchange Correlation Potential and Polarization Obtained from X-ray Wavefunction Refinement 107
4.1.1 Electron Correlation (EC) 109
4.1.2 Exchange Correlation Potential 114
4.1.3 Polarization 118
4.1.4 Derivation of the Effects from Experimental Data 122
4.1.5 Discussion of Results 125
4.2 *Publication:* Chemical Bonding in Polarised Push-Pull Ethylenes 126
4.3 *Publication:* Proximity Enforced Agostic Interactions Involving Closed-Shell Coinage Metal Ions 133
4.4 *Publication:* Fast and Accurate Quantum Crystallography: from Small to Large, from Light to Heavy 141

5 Applying Quantum Crystallography for Rational Drug Design 153
5.1 *Publication:* Similarities and Differences between Crystal and Enzyme Environmental Effects on the Electron Density of Drug Molecules 153
5.2 Sila-Ibuprofen 168
5.2.1 Synthesis 168
5.2.2 Characterization 169
5.2.2.1 Melting Enthalpy 169
5.2.2.2 Solubility 169
5.2.2.3 NMR Spectra 170
5.2.2.4 Stability in Aqueous Solution 175
5.2.3 Crystallography of Sila-Ibuprofen 176
5.2.4 Complementary Bonding Analysis 177
5.2.5 Electrostatic Potential 178
5.2.6 Force Field Development 180
5.2.7 MD Settings 186
5.2.8 MD Simulations 186
5.2.9 aNCI 188
5.2.10 Averaged Interaction Energies (aIE) 190
5.2.11 Free Energy Perturbation Calculations 198
5.2.12 Measurement of IC $_{50}$ 200
5.2.13 Toxicological Profile 201
5.2.14 *Publication:* Sila-Ibuprofen 201
5.3 Crystal and Enzyme Environmental Effect on Ibuprofen and Sila-Ibuprofen 211
5.3.1 Separation of Enantiomers & Co-Crystalization with Arginine 211
5.3.2 Comparison of Environments (Gas, Solvation, Crystal, Protein) 216
5.3.2.1 aNCI 217
5.3.2.2 Interaction Density 220
5.3.2.3 Interaction Electrostatic Potential 221
5.3.2.4 Bond-Centered Difference Density 224

6 Summary and Conclusions 233

7 Perspectives 236

Acknowledgements 238

Appendices 240

A Software Reference 242
- A.1 AIMAll . . . 242
- A.2 Crystal . . . 242
- A.3 CrystalExplorer . . . 242
- A.4 DGrid . . . 242
- A.5 Gaussian09 . . . 242
- A.6 Multiwfn . . . 242
- A.7 NAMD2 . . . 242
- A.8 NBO . . . 243
- A.9 Olex2 . . . 243
- A.10 ORCA . . . 243
- A.11 Tonto . . . 243
- A.12 VMD . . . 243

B Enabling Quantum Crystallography 244
- B.1 Supporting Information for Publication: Generalizing non-spherical structure refinement: Hirshfeld Atom Refinement in NoSpherA2 . . . 244
- B.2 Averaged Interaction Energy Scripts . . . 287

C Applying Quantum Crystallography for Complementary Bonding Analysis 290
- C.1 Supporting Information for *Publication:* Chemical Bonding in Polarised Push-Pull Ethylenes . 290

D Applying Quantum Crystallography for Rational Drug Design 322
- D.1 Supplementary Material for *Publication:* Similarities and Differences between Crystal and Enzyme Environmental Effects on the Electron Density of Drug Molecules . . . 322
- D.2 Supplementary Material for *Publication:* Sila-Ibuprofen . . . 331

Bibliography 364

Declaration of Consent 377

Curriculum Vitae 378

List of Abbreviations

A | B | C | D | E | F | G | H | I | L | M | N | P | Q | R | S | W | X

A

a.u. Atomic Units
ADP Atomic Displacement Parameter
aNCI averaged Non-covalent Interaction
AO Atomic Orbital

B

BCP Bond Critical Point

C

CIF Crystallographic Information File
COSX Chain of Spheres
COVID-19 Coronavirus Disease 2019
COX-I Cyclooxygenase I
COX-II Cyclooxygenase II
CPU Central Processing Unit
CSD Cambridge Structural Database

D

DFT Density Functional Theory
DI Delocalization Index
DKH Douglas Kroll Hess
DSC Differential Scanning Calorimetry

E

EC Electron Correlation
ECP Effective Core Potential
ED Electron Density
ELF Electron Localization Function
ELI Electron Localizability Indicator
ESP Electrostatic Potential

F

FEP Free Energy Perturbation

G

GGA Generalized Gradient Approximation
GPU Graphical Processing Unit
GUI Graphical User Interface

H

HAR Hirshfeld Atom Refinement
HF Hartree-Fock
HIV/AIDS Human immunodeficiency virus infection and acquired immune deficiency syndrome
HPLC-UV High performance liquid chromatography-ultra violet detection
HTS High-Throughput-Screening

I

IAM Independent Atom Model

L

LCAO-MO Linear combination of Atomic Orbitals to Molecular Orbitals
LDA Local Density Approximation
LS Least-squares

M

MD Molecular Dynamics
MM Molecular Mechanics
MO Molecular Orbital

N

NBO Natural Bond Orbital
NCI Non-covalent Interaction
NLMO Natural Localized Molecular Orbitals
NMR Nuclear Magnetic Resonance
NoSpherA2 Non-spherical atoms in Olex2
NPA Natural Population Analysis
NRT Natural Resonance Theory
NSAID Non-steroidal anti-inflammatory drug

P

PBC Periodic Boundary Condition
PDB Protein Data Base
PES Potential Energy Surface

PR Percentage Reconstruction

Q

QCr Quantum Crystallography
QM Quantum Mechanics
QM/MM Quantum Mechanics Molecule Mechanics hybrid method
QTAIM Quantum Theory of Atoms in Molecules

R

RGBI Roby-Gould Bond Index
RI Resolution of Identity
RMSD Root Mean Square Displacement
RRS Real-Space R-Value

S

SCF Self-Consistent field

W

WHO World Health Organization

X

XCW X-ray Constrained Wavefunction fitting
XWR X-ray Wavefunction Refinement

List of Figures

1.1 Picture of a Diffractometer . . . 2
1.2 Isosurface of Electron Density from Fourier Synthesis . . . 5
1.3 Scattering factor for fluorine, fluoride and carbon . . . 6
1.4 Residual Density after IAM Refinement . . . 7
1.5 Comparison of spherical and non-spherical density . . . 8
1.6 Scattering factor for carbon in IAM and HAR . . . 13
1.7 Scattering factor for Hydrogen in IAM and HAR . . . 14
1.8 Schematic: Key-Lock Principle . . . 19
1.9 Visualisation of the *umpolung* in Trimethylsilane . . . 26
1.10 Scheme of covalent bond formation by E64c drug molecule in enzyme . . . 27
1.11 Ibuprofen Arginine hydrogen bonds . . . 27

2.1 Scheme of a 4-circle Diffractometer . . . 29
2.2 Pictures of Diffractometer with Labels . . . 30
2.3 Scheme of Ewald construction . . . 32
2.4 Plot of χ^2 against λ for Theoretical Example Data . . . 39
2.5 Qualitative scheme of Fermi- and Coulomb-Hole . . . 46
2.6 Plot of Interaction Density for Water . . . 53
2.7 Plot of NCI for a Tetramer of Water . . . 54
2.8 Plot of NCI and aNCI for Ibuprofen in COX-II for 1 Frame and 1000 . . . 56
2.9 Definition of Force Field Parameters . . . 62
2.10 Non bonded Energies against λ in FEP Simulations . . . 66

3.1 Scheme of Interfaces between Software in NoSpherA2 . . . 71
3.2 Scheme of individual steps in NoSpherA2 . . . 72
3.3 Structure of THPP after HAR with Labels . . . 74
3.4 Plot of Difference in *.tsc* files for Disorder . . . 75
3.5 Structure of L-Alanine and Ammonium Borate used for benchmarks . . . 76
3.6 Plot of Structure Factor Agreement based on Chosen Basis Set . . . 77
3.7 Water arrangement for averaged Interaction Energy example . . . 97
3.8 Average Interaction Energy and Contributions example . . . 98
3.9 Bond-wise Cube Alignment . . . 103

4.1 Difference between CCSD and HF wavefunction ED . . . 109
4.2 Difference in ED during XCW against CCSD (0.7 Å^{-1}) . . . 110
4.3 Difference in ED during XCW against CCSD (1.4415 Å^{-1}) . . . 111
4.4 Difference in ED during XCW against CCSD (2.0 Å^{-1}) . . . 112
4.5 Graphs of structure factor and ED agreement of XCWs for CCSD . . . 113
4.6 Difference in EC Potential λ against λ_0 (0.7 Å^{-1}) . . . 115
4.7 Difference in EC Potential λ against λ_0 (1.4415 Å^{-1}) . . . 116
4.8 Difference in EC Potential λ against λ_0 (2.0 Å^{-1}) . . . 117
4.9 Difference between periodic and HF wavefunction ED . . . 118

4.10 Difference in EC Potential λ against λ_0 (0.7 Å^{-1}) 119
4.11 Difference in EC Potential λ against λ_0 (1.4415 Å^{-1}) 120
4.12 Difference in EC Potential λ against λ_0 (2.0 Å^{-1}) 121
4.13 Graphs of structure factor and ED agreement of XCWs for CRYSTAL calculations 122
4.14 Difference Density from experimental XCWs at 0.7 Å^{-1} 123
4.15 Difference Density from experimental XCWs at 1.4415 Å^{-1} 124

5.1 Scheme of Synthesis of sila-ibuprofen 169
5.2 Calibration of Solubility Measurement 170
5.3 ^{1}H-NMR of sila-ibuprofen 170
5.4 ^{13}C{^{1}H}-NMR of sila-ibuprofen 171
5.5 ^{29}Si{^{1}H}-NMR of sila-ibuprofen 171
5.6 ^{1}H-NMR of hydroxy-sila-ibuprofen 172
5.7 ^{13}C{^{1}H}-NMR of hydroxy-sila-ibuprofen 172
5.8 ^{29}Si{^{1}H}-NMR of hydroxy-sila-ibuprofen 173
5.9 ^{1}H-NMR of sila-ibuprofen-siloxan 174
5.10 ^{29}Si{^{1}H}-NMR of sila-ibuprofen-siloxan 174
5.11 Stability of sila-ibuprofen in Aqueous Solutions 175
5.12 Complementary bonding analysis on XWR of ibuprofen and sila-ibuprofen 178
5.13 Hirshfeld-Surfaces with mapped ESP 179
5.14 Missing Force Field Parameters 180
5.15 Performance of Dihedral Force Field parameters 181
5.16 Model Compound for Lennard-Jones Parameters 181
5.17 Performance of Lennard-Jones Parameters 182
5.18 Ibuprofen/sila-ibuprofen arginine Lewis Structure 186
5.19 Distances between ibuprofen/sila-ibuprofen and arginine in COX-I 187
5.20 Distances between ibuprofen/sila-ibuprofen and arginine in COX-II 188
5.21 aNCI Plots of ibuprofen and sila-ibuprofen in COX-I 189
5.22 aNCI Plots of ibuprofen and sila-ibuprofen in COX-II 189
5.23 Interaction Energies of ibuprofen in COX-I 191
5.24 Total Interaction Energy of ibuprofen in COX-I 191
5.25 Interaction Energies of sila-ibuprofen in COX-I 192
5.26 Total Interaction Energy of sila-ibuprofen in COX-I 192
5.27 Interaction Energies of ibuprofen in COX-II 195
5.28 Total Interaction Energy of ibuprofen in COX-II 195
5.29 Interaction Energies of sila-ibuprofen in COX-II 196
5.30 Total Interaction Energy of sila-ibuprofen in COX-II 196
5.31 FEP Simulation Residue 199
5.32 Plot of Free Energy Calculated in COX-I 199
5.33 Plot of Free Energy Calculated in COX-II 199
5.34 Results of *in-vitro* Measurement of IC $_{50}$ Values 200
5.35 Refined Structure of ibuprofen and sila-ibuprofen - PEA Co-Crystals 212
5.36 Structure of Ibuprofen-Arginine Co-crystal 214
5.37 Structure of Ibuprofen-PEA simulation Box 216
5.38 aNCI Plots of ibuprofen and sila-ibuprofen in the PEA co-crystals 219
5.39 Interaction Density of ibuprofen and sila-ibuprofen 221
5.40 Interaction Electrostatic Potential of ibuprofen and sila-ibuprofen 223
5.41 Bond Labeling Scheme for Bond-centered Analysis 224
5.42 Bar-plot of Cumulative Difference Density for ibuprofen and sila-ibuprofen 225
5.43 Bar-plot of Cumulative R $_{RS}$ Value for ibuprofen and sila-ibuprofen 225

5.44 Bar-plot of Cumulative Integrated Electron Sum for bonds around exchange position in ibuprofen and sila-ibuprofen 227
5.45 Bar-plot of Cumulative R $_{RS}$ Value for bonds around exchange position in ibuprofen and sila-ibuprofen 227
5.46 Bar-plot of Cumulative Integrated Electron Sum for Bonds in the vicinity of carbon/silicon exchange in ibuprofen and sila-ibuprofen 228
5.47 Bar-plot of Cumulative R $_{RS}$ Value for Bonds in the vicinity of carbon/silicon exchange in ibuprofen and sila-ibuprofen 228
5.48 Bar-plot of Cumulative Integrated Electron Sum for Common Bonds in ibuprofen and sila-ibuprofen 229
5.49 Bar-plot of Cumulative R $_{RS}$ Value for Common Bonds in ibuprofen and sila-ibuprofen 230

List of Tables

4.1 Table of fitting parameters for reconstruction of electron correlation density of XCW 114
4.2 Table of fitting parameters for reconstruction of interaction density of XCW 121

5.1 Refinement Statistics of ibuprofen and sila-ibuprofen 177
5.2 Politzer Parameters of Ibuprofen and Sila-Ibuprofen 179
5.3 Force Field: Bonds and Angles 183
5.4 Force Field: Dihedral and improper 184
5.5 Force Field: Lennard Jones and Charge 185
5.6 Averaged Interaction Energies in COX-I 193
5.7 Averaged Interaction Energies in COX-II 197
5.8 Refinement Statistics of ibuprofen and sila-ibuprofen - PEA Co-Crystals 213
5.9 Refinement Statistics of ibuprofen arginine Co-crystals 215
5.10 Force Field for PEA: Bonds, Angles and Dihedral 217
5.11 Force Field for PEA: Lennard Jones and Charge 218
5.12 Difference Measures for the Interaction Density of ibuprofen and sila-ibuprofen 220
5.13 Difference Measure for the Interaction Electrostatic Potential of ibuprofen and sila-ibuprofen . 222
5.14 QTAIM Charges of Ibuprofen and Sila-Ibuprofen in different environments 226

Abstract

The field of crystallography is a key branch of natural sciences, important not only for physics, geology, biology or chemistry, but it also provides crucial information for life sciences and materials science. It laid the foundations of our textbook knowledge of matter in general. In this thesis, the field of quantum crystallography – a synergistic approach of crystallography and quantum mechanics – is used as a tool to predict and understand processes of molecules and their interactions. New methods are proposed and used that provide deeper insight into the influence of local molecular environments on molecules and allows advanced predictions of the biochemical effect of drugs. Ultimately, this means that we can now understand interactions between molecules in crystal structures more completely that were long thought to be fully characterized.

As part of this work, new software was developed to handle theoretical simulations as well as experimental data – and also both of them together at the same time. The introduction of non-spherical refinements in standard software for crystallography opens the field of quantum crystallography to a wide audience and will hopefully strengthen the mutual ground between experimentalists and theoreticians. Specifically, we created a new native interface between Olex2 and non-spherical refinement techniques, which we called NoSpherA2. This interface has been designed in such a way that it can be used for any kind of non-spherical atom descriptions. This will allow refinement of modern diffraction data employing modern quantum crystallographic models, leaving behind the century old Independent Atom Model (IAM).

New software was also developed to provide novel models and descriptors for understanding environmental effects on the electron density and electrostatic potential of a molecule. This so-called Quantum Crystallographic Toolbox (QCrT) provides a framework for the fast and easy implementation of various methods and descriptors. File conversion tools allow the interfacing with many existing software packages and might provide useful information for future method development, experimental setups and data evaluation, as well as chemical insight into intra- and intermolecular interactions. It is fully parallelized and portable to graphic card processors (GPUs), which provide extraordinary amounts of computational power with moderate resource requirements. Especially in the context of ultra-bright X-ray sources like X-ray free electron lasers and electron diffraction these new models become crucial to have a better description of experimental findings.

In applying this new framework of quantum crystallographic methods, we analyze a type of bonding at the edge of conventional organic chemistry: The push-pull systems of ethylenes. We show how X-ray constrained bonding analysis leads to the unambiguous determination of the behavior and type of bonding present in a series of compounds which are contradicting the Lewis-picture of a double-bond.

This new understanding has led to the development of a new potential drug, namely a silicon analogue of ibuprofen; one of the most important drugs known to humankind. We determined its physical properties, investigated its stability and potency as a more soluble and novel alternative of ibuprofen: While retaining the same pharmaceutical activity of ibuprofen, making it a bioisoster for ibuprofen, this material shows a better applicability in aqueous media.

Chapter 1

Introduction

1.1 Brief History and Introduction to Crystallography

To provide context to the benefits of quantum crystallography – and the requirements of the related techniques – a small outline of central aspects of crystallography and its application will be given in this introduction. Since the field is very broad and has manifold applications the introduction will aim towards the developments of quantum crystallography in the context of wavefunction-based techniques and the distinction from previous approaches, while being aware that this is only a part of the field and not claiming completeness of the topic in general.

1.1.1 Crystallography

Without a doubt, crystallography is a fundamental branch of science that is closely linked with advances in fields not only limited to chemistry, physics, geology and biology: no less than 25 Nobel Prizes have been awarded in the context of crystallography since 1900. [1] Some selected breakthroughs in science were achieved using crystallography and crystallographic methods including the description of the DNA-double helix structure, [2] the huge number of databased protein structures [3] based on protein crystallography, the discovery of graphene [4] or the discovery of quasi-crystals, [5] which ultimately led to a new formulation of the definition of a crystal. The current definition by the International Union of Crystallography describes a crystal as matter that shows concentrated *Bragg* diffraction, besides the always present diffuse scattering. [6]

1.1.2 Diffraction Experiment

The diffraction experiment is a setup where at least one crystal is positioned inside a beam of photons (X- or γ-ray), neutron or electrons, that is subsequently scattered by the crystal due to its arrangement of atoms that can be understood as a lattice. This diffracted pattern is recorded using detectors which can not only qualitatively, but quantitatively measure the intensity of each diffraction maximum, the Bragg peaks. This diffraction pattern is analyzed in terms of position and integral of each diffracted beam. Using the selection rules for three, or in the case of quasi-crystals higher, dimensional lattices and orientation of the diffraction pattern the lattice type is determined. The first observation of diffraction by crystals was reported by *Max von Laue*, [7] the diffracted pattern explained mathematically in detail by *William Henry* and *William Lawrence Bragg*. [8–10] It can be interpreted in terms of a reciprocal lattice through the construction of the so-called *Ewald*-sphere, [11] that allows the prediction of the orientation of a crystal necessary to be in *Bragg* condition to observe a diffraction peak. This technique, as e.g. used in a *Debye-Scherrer* [12] arrangement, where the superposition of all peaks is recorded on a single image through rotation of the crystal, led to first measurements and interpretations of diffraction data.

These new techniques rendered applications to fields ranging from the analysis of small molecule liquid phases to the determination of macromolecular protein structures possible. [13–15] The recorded diffraction pattern can be used to create and refine a model of the arrangement, motion and shape of atoms in the crystal, which will be discussed in more detail in the following sections and chapter 2. A single crystal diffractometer, as used at the

Department of Chemistry and Biochemistry of the University of Bern, is shown in Figure 1.1. A more detailed description of the parts and setup of the diffractometer axes will be given in section 2.1.2. Since the diffraction experiments in this thesis were all single-crystal X-ray diffraction experiments only this type of diffractometer will be discussed in detail, leaving aside the field of powder diffraction.

Figure 1.1: The single-crystal X-ray diffractometer used in the Department for Chemistry and Biochemistry at the University of Bern for crystal structure service and high-quality high-resolution measurements.

1.1.3 Radiation sources

Diffraction experiments can be performed with various types of radiation. Photon radiation of various wavelengths can be used, for example, X-ray or γ-ray. The latter is, due to the difficultly of generating the radiation, less prominent, but successful diffraction experiments are reported in the literature. [16] X-ray beams are much easier to produce and also offer a wide range of wavelengths, depending on the source of the photons. Another type of radiation used to perform diffraction experiments is neutron radiation. In contrast to photon radiation, neutron radiation scatters due to the interaction with the magnetic moment of the nucleus of atoms. An example of experiments with neutron radiation is the work on heavy ice. [17] The photons of X- and γ-radiation mainly interact with the electrons present in the crystal.

Recently, an increase of accuracy and experimental interpretation of diffraction experiments using electron radiation promises an additional radiation type for high-quality diffraction experiments, while electron diffraction was already performed for a long time. [18] Electron radiation is diffracted by the electrostatic potential present in the crystal, which is a combination of nuclear charges and electron density effects. The size of particles required for diffraction experiments is much smaller, which allows, for example, the analysis of surface compositions and structures of crystals. [19]

In this thesis, the radiation used and the interpretation methods developed are focused on the use of X-ray diffraction. But in principle, except for neutron radiation, all other types of radiation can also be used when employing the correct transformation to the calculated models. This will be discussed after the presentation of new methods during this thesis in the perspectives chapter (chapter 7).

In-house X-ray sources

Modern lab X-ray sources usually are limited to fixed wavelengths, since the origin of the radiation is in most cases copper, silver or molybdenum which is targeted with an electron beam to emit photons with characteristic wavelengths in addition to broader Bremsstrahlung. Usually, the desired element-characteristic monochromatic radiation is achieved by the use of filters and mirrors, acting as monochromators. [20, 21] This radiation is, however, only of limited brilliance, as the physical effect of photon emission is limited to the surface area targeted by the electron beam and the electric current available for excitation. With modern in-house µ-focus sources, a dataset of sufficient quality for quantum-crystallographic modeling can be obtained for small molecules with sufficient crystal quality, as shown in this thesis (see for example section 4.2).

Synchrotron radiation

In most cases the crystal quality is not as ideal, making high-quality and high-resolution investigations more difficult. Therefore, a brighter, stronger X-ray source might make Quantum crystallographic experiments possible even with limited scattering power of a crystal. One possible source of high intensity, high brilliance and high energy X-rays is a synchrotron. In contrast to the conventional X-ray sources using a metal target, these synchrotron sources accelerate electrons to near-light-speed using strong electromagnetic fields and force these electrons to fly in a circular arrangement of magnets keeping them focused and on track. This acceleration of the electrons in a circular trajectory causes them to emit a continuous spectrum of electromagnetic radiation, the so-called synchrotron radiation. This radiation can be used for various purposes, including diffraction experiments, where a certain wavelength of photons is selected through monochromators and mirrors and focuses on the position of a goniometer, similar to the apparatus of an in-house diffractometer, basically exchanging the conventional X-ray source with the large-scale synchrotron-source. [20]

Usually, preselected crystals are taken to a third-generation synchrotron with high brilliance and variable wavelength. These facilities provide high-quality datasets in a shorter time and with higher resolution, given sufficient crystal quality. Unfortunately, the number of available facilities of this kind is limited and not as easily accessible and orders of magnitude more expensive, concerning operation and setup, in comparison to conventional X-ray sources. Fortunately it was possible to conduct some of the diffraction experiments presented in this work at a synchrotron beamline. The studies presented in sections 5.2.14 and 5.3 are based on the diffraction experiments carried out at beamline BL02B1 at SPring-8 in Hyogo, Japan.

1.1.4 Detectors

The detection of X-rays is crucial for the conduction of diffraction experiments. At the beginning of history of diffraction experiments, the use of ionization chambers or point detectors, that were moved to measure the individual diffraction peaks by alignment of the crystal and detector, was the state-of-the-art methods, yielding high-quality quantitative diffraction data. [20] The use of photographic films to record the intensity of the diffracted beam could detect many reflections at the same time, which reduced the time needed to perform a diffraction experiment, but the estimation of blackening of the film was not as precise as the use of ionization counters. The benefits of the simultaneous detection of many reflections available from area detection were combined with high accuracy and precision of intensities through the development of Charge-Coupled-Device (CCD) detectors. They use the same kind of chip present in modern digital cameras combined with a layer of fluorescence material, for example, gadolinium oxide sulfide. This material lights up when irradiated with X-rays and the intensity of this light, which has lower energy and therefore is easier to detect, is evaluated by the CCD-chip. [21]

Especially modern hybrid pixel detectors with novel detection materials like CdTe made a switch of paradigm possible: Instead of trying to use as much of the measurement time at a radiation source by employing simple but fast measurement strategies, the speed and sensibility of the new pixel detectors now allows the careful planning of experiments and re-collection of some parts of the datasets while the crystal is still on the diffractometer. These detectors use highly X-ray sensitive diodes in combination with individual readout chips (a hybrid of these two building parts for each pixel, hence the name) to measure extremely precisely the intensity of diffraction patterns in an area detector, especially since the pixel detectors have much lower noise levels compared to CCD-detectors. [22] The detection is so fast and sensible, that high-quality experiments can be performed within a much shorter time. Older, slower detectors allowed only a highly limited number of measurements within a beamtime slot at a synchrotron facility. These advances in detector techniques and high-intensity radiation sources allow the experiment to be planned thoroughly and conducted as required to yield the highest possible number and quality of diffraction data. The biggest drawback of the newest kind of detectors is a high cost in production and limited time, in which the detector material, in case of CdTe, is stable, as it decomposes slowly. [23] In the course of this thesis a broad variety of detectors was used. The experiments performed by me in section 3.1.3 were done using a Hybrid-Pixel detector. The measurements in section 4.2 were performed on a CCD detector, while the synchrotron experiments in sections 5.2.14 and 5.3 were conducted using a curved image plate.

1.1.5 Structure Solution and Refinement

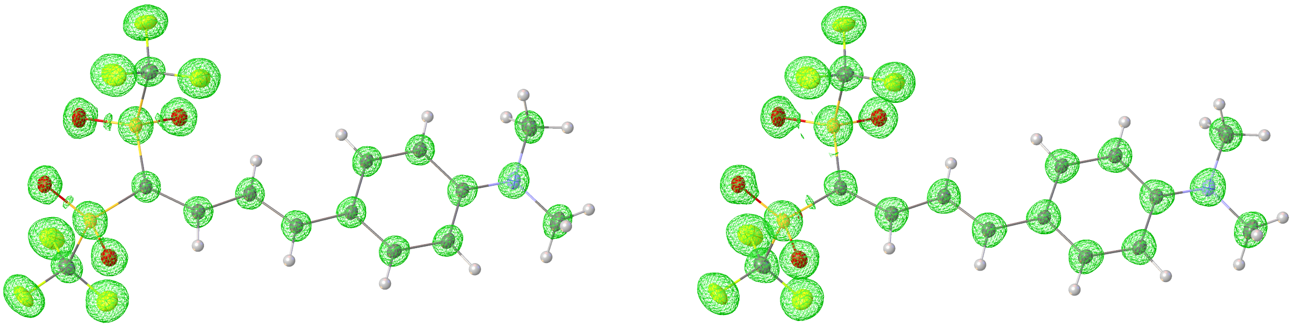
The measured diffraction pattern can be used to determine the three dimensional structure of the molecules present in the crystal used for the experiment. This structure solution is performed taking into account the information on the position of diffraction maxima and their intensity to obtain a first guess of the structure at hand. After an initial solution the model can be improved and finalized during the refinement.

The principle of structural refinement is based on the optimization of agreement between a model diffraction pattern of the structure under investigation and the measured diffraction intensities obtained from a conducted diffraction experiment of a crystal of that substance. [20,24] In all methods, the main procedure is the calculation of differences between the model and the experimental data and their subsequent minimization using a least-squares approach, as initially proposed by *Legendre*. [25, 26] The positions, shape of atoms and/or their displacement parameters are variables to maximize the agreement. In most cases, the shape of an atom is kept fixed throughout the least-squares procedure, but in principle could also be used as a fitting parameter. Besides the least-squared procedure is performed on weighted measurement data, to account for experimental errors, which however introduces a series of statistical biases. [27] This established mathematical framework allows the use of very well defined statistical tools and agreement statistics like R-factors, normal distributions or goodness of fit, which are mathematically indifferent of the model used to produce the calculated data. [28–30] In this thesis the focus is drawn to the conduction of X-ray diffraction experiments and their interpretation. Since the diffraction of the photons of the incident X-ray beam is based on the interaction of electrons with the photons, the intensity of the measured pattern is the Fourier transform of the spatial Electron Density (ED) distribution in the crystal. [20] An adequate theory needs to be selected for the description of the ED of the structure under investigation to obtain a model to be used in the optimization of agreement with the measured pattern to describe the measured intensities correctly with physically reasonable assumptions.

1.1.6 X-ray Data Quality and Resolution

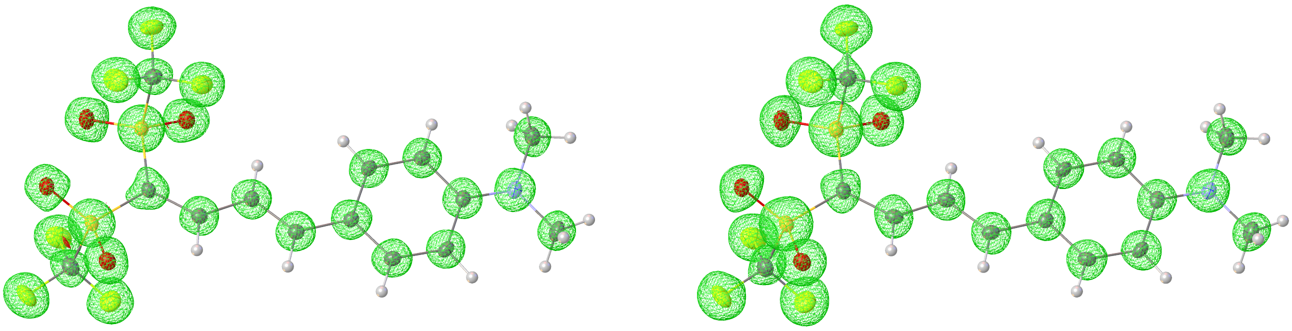
A very important point of interest in modern X-ray diffraction experiments and their interpretation is the issue of data quality and maximum achievable resolution. Data quality refers to the internal agreement and measurement uncertainty of a dataset. If a peak in the diffraction pattern is measured multiple times, the agreement between these occurrences is a measure of data quality, as well as the uncertainty of the integrated intensity of a peak. It is estimated by the level of noise in the data, for example taking into account the size of background intensity measured by a detector in regions where no diffraction peak is found. If the data quality is low, the uncertainty associated with measured intensities becomes high and therefore the significance, that is, the ratio of signal-to-noise becomes low. This will result in poor precision of the resulting structural model, as the certainty of a fitted parameter, for example, the position of an atom, will become very low as a result.

The maximum achievable resolution of a diffraction experiment is of importance, as the ED that is observed resulting from the experimental data is very sensitive to the number of observed data points. This is a logical consequence of the relation between the real space electron density and the diffraction pattern in reciprocal space. A Fourier series to calculate the ED based on observed diffraction data is truncated at the maximum resolution of the experimental data. The ED derived by this series is the property used in X-ray diffraction experiments to model the structure, therefore the incompleteness of the series will directly affect the quality of the model. Since an early truncation of a Fourier series can introduce significant inaccuracies, the highest resolution, corresponding to the latest possible truncation of the Fourier series, is desirable. The resolution of the diffraction pattern, which is proportional to the angle θ between the primary beam and the diffracted beam, directly influences what features of the density calculated from a Fourier series are distinguishable. At low resolution (high d_{min}) the distinction of individual atoms already becomes difficult, while at high resolution even maxima on the bond between two atoms or in the region of lone pairs might be observed and distinguishable from the atomic core. An example of a plot of the evolution of the effect of different resolutions on the calculated ED from experimental diffraction patterns is shown in Figure 1.2.



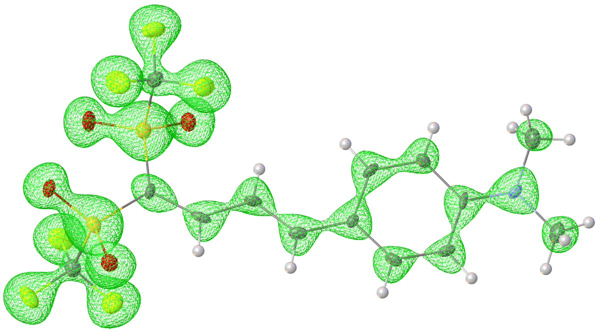
(a) d_{min} = 0.45 Å

(b) d_{min} = 0.6 Å



(c) d_{min} = 0.8 Å

(d) d_{min} = 1.0 Å



(e) d_{min} = 1.2 Å

Figure 1.2: Electron Density from Fourier synthesis (compare equations 2.6 and 2.3) of experimental data with different resolution (in d-spacing; lower d-spacing is better resolution), obtained by truncating the synthesis at different cut-off values, marked below each subfigure. Visualized using Olex2. [31] Electron density isosurface mesh in green. Isovalue = 2.9 eÅ^{-3}. Dataset used is one of the molecules of the study "Chemical Bonding in Polarized Push-Pull Ethylenes", as shown in Section 4.2.

The data used to produce Figure 1.2 was high-resolution data with low noise and good statistical agreement. If the scattering power of the crystal already became insignificant at, for example, 1.0 Å the noise would be much bigger and introduce not only Fourier truncation problems, but also the individual signals would become more problematic. This is a problem the field of protein crystallography faces quite often, as their highest obtainable resolution often is lower than 1.0 Å. A check of the Protein Data Base (PDB) shows only around 0.5% of all deposited structures were measured with resolution higher than 1.0 Å. [3]

Additional problems

In the case of disordered molecules or solvent in the crystal structure, it is often difficult to obtain high quality and high-resolution data. Even if crystals with disorder and solvent molecules can be measured to sufficient data quality, the refinement becomes more challenging. Disordered atoms and partial occupancies, make refinements more complicated and less robust, as there are much more parameters and correlations between variables during the refinement. It is possible to remove density of atoms that are too difficult to model based on the analysis of voids and their contained density based on the diffraction pattern to circumvent these problems, for example using *SQUEEZE* or *BYPASS*. [32, 33]

1.1.7 Independent Atom Model (IAM)

The first model to describe measured intensities from X-ray diffraction, either obtained by single crystal or powder diffraction experiments, was the Independent Atom Model (IAM). [34–40]

Promolecule Atoms

In this model, a spherical ED distribution of promolecular, that is non-interacting, atoms is assumed. This reduces the problem of determining the contribution to the ED at a given point in space by an atom to the sole evaluation of the distance to the position of its nucleus. This allows for a simple tabulation of atomic form factors, also called scattering factors, f. These are the Fourier transforms of the electron density of an atom with respect to the reciprocal space, which are precomputed using atomic wavefunctions. [41] These f are ready to be used during a structural refinement without additional calculations. An example of these scattering factors with respect to the resolution in d-spacing of the diffraction pattern is given in Figure 1.3.

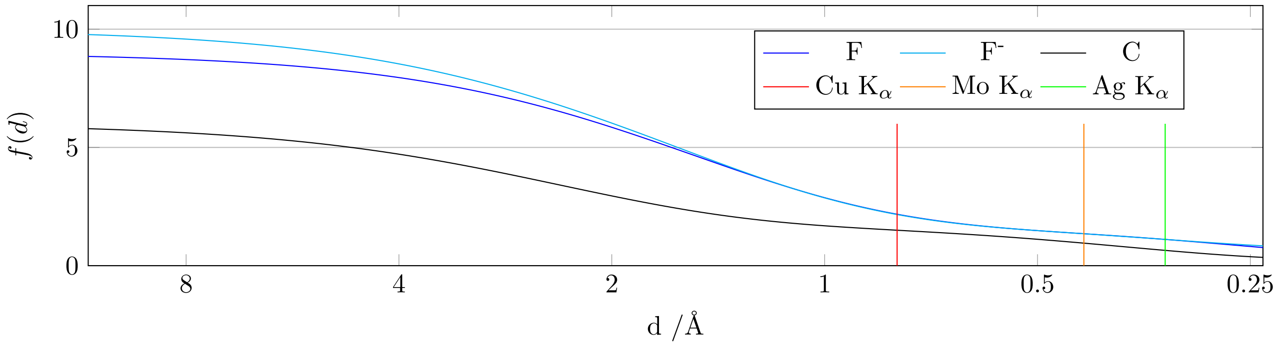


Figure 1.3: Plot of the value of the scattering factor with respect to d-spacing for the neutral fluorine atom, the fluoride anion and the neutral carbon atom. Values of the exponential functions used for plotting as printed in the International Tables of crystallography. [41] Colored lines representing the diffraction limit of common radiation types are added to visualize the importance of differences using each radiation type.

This tabulation makes these refinements almost instantaneous for small molecules on normal modern desktop computers. It is noteworthy that the scattering factor between neutral and single negatively charged fluorine, as shown in Figure 1.3, is mostly different in the low-resolution region (high d values). At higher resolution the difference is minimal. Limits of resolutions reachable using the corresponding type of in-house radiation are indicated to show that the charge difference between fluorine and fluoride can already be observed using copper radiation. This would imply that the distinction of a partial charge transfer could also already be observed based on a diffraction experiment using any type of the most common radiation types. The significant downside of this IAM is the complete neglect of the non-spherical shape and behaviour of atoms in molecules, disregarding bonds, lone pairs and basically everything which goes beyond the atomic models of *Rutherford* or *Bohr*. [42–45]

Shortcomings

Modern single-crystal X-ray diffraction experiments and even high-quality powder diffraction experiments [46,47] can show good quality data at higher resolution in reciprocal space compared to any setups available at the time when the IAM was introduced. The information about the non-spherical character of molecules contained in the measured data is neglected by applying this model and can therefore not explain all measured features of the data, by restriction of the model used. But also 50 years ago the missing information of covalent bonding and polarization of atoms was already discussed in the literature and attempts to model them were done. [48,49] The extent, spatial distribution and significance of this missing information on bond density in the model of the structure is shown exemplarily in Figure 1.4.

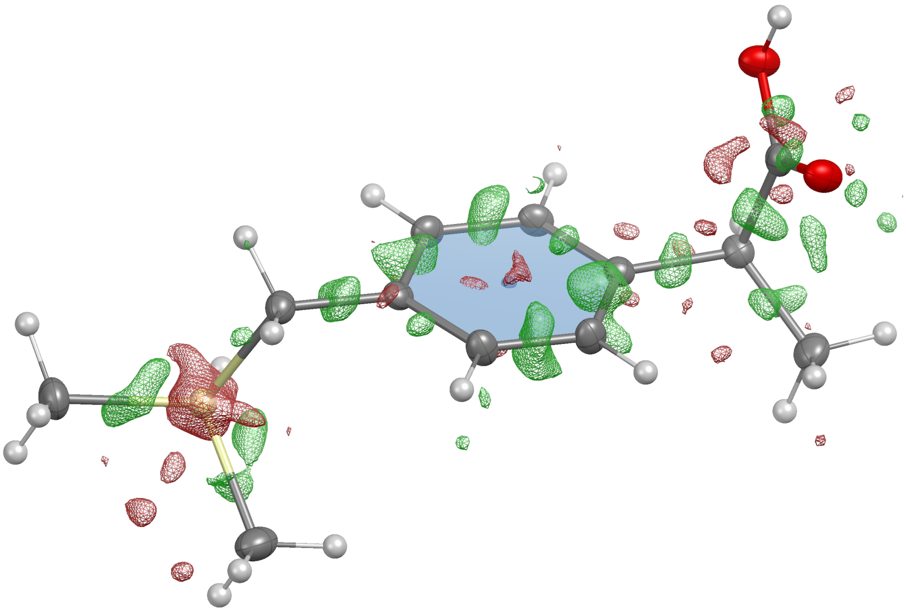


Figure 1.4: Residual electron density after IAM refinement using olex2.refine, visualized using Olex2. [31,50] Positive residual density in green, negative residual density in red. Ring plane visualized in blue. Isovalue +/- 0.3 $e\text{Å}^{-3}$. The dataset used for this figure is the measurement of sila-ibuprofen as presented in section 5.2.

The electron density in regions where covalent bonding is expected is clearly underestimated, especially when a high double bond character or lone pairs are present coinciding with an overestimation of electron density in regions of rings or above and below the plane of sp^2 hybridized atoms, as seen for the carboxylic acid carbon atom. It should be highlighted that using the IAM one can accurately obtain the position and Atomic Displacement Parameters (ADPs) of non-hydrogen atoms from even low quality diffraction data, since the chemical connectivity could be deduced from distances by comparison among series of different structures. While there might be a correlation of bond order and the distance between two atoms, this correlation is purely empirical. [51–53] There is no possibility to perform chemical analysis on the bonding situation of atoms by solely employing the IAM.

Another way to look at the model differences and shortcoming is to look at the electron densities obtained by the IAM and a theoretical wavefunction calculation of benzene side by side (for details on wavefunction calculations see section 2.2). Isosurfaces of the electron density of the IAM and a molecular wavefunction are shown in Figure 1.5. The superposition of the spherical electron densities of the IAM on the left show a deformation of the isosurfaces from spherical shape, but at the highest isovalue, which is the innermost surface, the connection between atoms is broken. This is not the case in the electron densities derived from the wavefunction. All surfaces form evenly shaped envelopes around the bonds and also hydrogen atoms are not separated from the rest of the molecule at higher isovalue.

If the density of the IAM is compared to the shape of the 6-membered ring in the molecule in Figure 1.2 it might show why the IAM was so successful and sufficient for such a long time: the shape of the observed density at the resolution of 1.2 Å resembles that of the IAM quite well. And the average resolution of the diffraction experiments only increased over time, as the experimental setup became more sophisticated and the cooling of crystals became more common. Also the limitation in computational power made the use of more sophisticated models unfeasible for a long time.

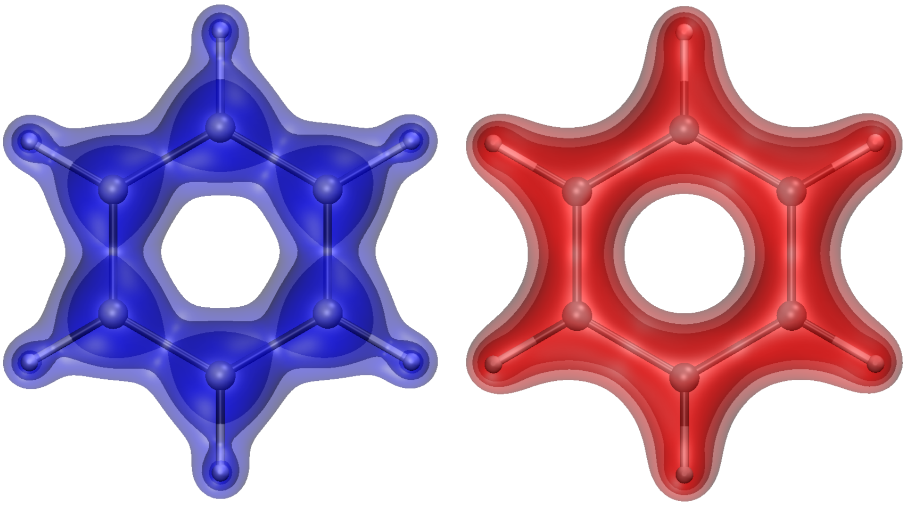


Figure 1.5: Isosurfaces of the electron density at 0.12, 0.17 and 0.22 Atomic Units (a.u.) isovalue (with increasing opacity of the surface) for benzene using the IAM (left, blue) and from a wavefunction calculated using *ORCA* [PBE/cc-pVTZ] (right, red). [54,55] Electron densities calculated using *cuQCrT* (see section 3.3), plotted using VMD. [56]

Hydrogen atoms in the IAM

The most severe problem arises for hydrogen atoms. Since their only electron in most cases is involved in a covalent bond to a different element, often having a higher electronegativity, the assumption of a spherical behaviour of the atoms fails completely. This leads to significantly shortened bonds if the position of hydrogen atoms is not restricted in some way. Usually restrains or constraints are used, e.g. the so-called riding model. [57,58] These can only estimate the position of the atom by geometric constructions and their displacement parameter as a fixed multiple of the non-hydrogen atom it is attached to. This model cannot distinguish between, for example, electron pulling effects of elements with significantly higher electronegativity leading to a shift of the bond-density or the presence of a hydrogen bond elongating the H-X bond and withdrawing density from the bond. This is the origin of a common believe among chemists, that hydrogen atom positions *cannot* be measured with X-ray diffraction, as they are not "visible" enough. [59–65]

1.1.8 Non-Spherical Refinements - History and Models

The observed inaccuracies of the IAM necessitated the development of improved models using different approaches to tackle the problem of the missing description of the bond- and lone-pair-density. All of them can be summarized under the term non-spherical atom models, to distinguish them from the spherical IAM.

Multipole Model

The first and very successful model of a description based on non-spherical atoms is the Multipole Model. This model extends the structure factor equation from the IAM by using a series of spherical harmonics functions to account for asphericity and other neglected effects, like shell contractions, charge transfers in the valence shell etc. to provide a more sophisticated scattering factor for each atom. [66–72]

Improvements

The model was very successful in modelling non-spherical atoms and is described in the literature. [73–77] In this context, one has to distinguish between two types of refinement: (i) the refinement of the multipoles in conjunction with the refinement of positions and ADPs, which is a method to obtain on the one hand the electron density of the molecule and on the other hand the structure at the same time. (ii) The refinement of structural and displacement parameters, using a pre-calculated description of the non-spherical atomic shape. There are three batabanks of pre-defined atom types and correpsonding multipoles: the *Invariom* approach, [78] the ELMAM database [79] or the *UBDB*, which is newly branded with the name Multipolar Atom Types from Theory and Statistical clustering (MATTS). [80, 81] While (i) is a refinement of the charge density based on the diffraction data, (ii) is a pure structural refinement, identical in any way to the IAM, except for the use of more correct scattering factors. This type of refinement will be referred to as non-spherical structural refinement to distinguish it from the refinement of the density (i).

In most cases, it had been impossible to meaningfully describe e.g. the anharmonic motion of atoms, as the approximation of the IAM was too crude to filter residual bond and lone pair density from contributions from the anharmonic description of atomic movement. This has become feasible and reasonable within the applications of the Multipole Model. [82–85] In this way, 3rd and 4th order Gram-Chalier expansions for the anharmonic displacement of atoms were introduced into software like *XD* or *MoPro*. [86–88]

A major improvement compared to the IAM is the possible analysis of the obtained electron density distribution. Topological analyses, as well as calculations of derived properties like Electrostatic Potential (ESP), Laplacian and atomic polarizabilities, became possible. For further details about these properties please refer to section 2.3.

Shortcomings

A major drawback of this method is the dependence on the setup of local coordinate systems for the spherical harmonics and symmetry restrictions, especially in the case of transferable atoms from databases. This is overcome by applying the automations in mentioned database approaches, but only to a certain degree. The database only works for previously defined systems like proteins and amino acids, but mostly small-molecule crystallography is interested in novel bonding situations or elements, which are not part of these databases.

Another major drawback of the theory itself is the limitation of the basis set to a single ζ-slater type radial function (for details see section 2.2.2) used for the calculation of multipole scattering factors and especially the flexibility of very polar bonds, where features of the multipole model do not correspond to any observable trend in Quantum Mechanics (QM) reference calculations. [89–92] This might also be due to a certain ambiguity in the choice of multipole populations and their covariance, as structural differences have a strong impact on multipole parameters and *vice versa*.

Hydrogen atoms in the Multipole Model

This also holds for the problematic determination of hydrogen atoms, as in most cases the symmetry of the multipoles of hydrogen atoms are restricted to obey high symmetry constrains and therefore reduce to very limited freedom for a redistribution of the ED originating from the hydrogen atoms. This usually yields in high covariances of ED-parameters and positions, as well as usually difficulties to obtain a good anisotropic description of hydrogen ADPs. Recent findings report multipole-database approaches to be able to reproduce the bond lengths and thermal discplacement parameters of hydrogens observed by neutron diffraction better compared to the IAM. [93] But compared to other non-spherical structural refinement methods the results are still not ideal.

Maximum Entropy Methods

The maximum entropy method, as proposed by *Jaynes*, is a method to derive properties out of limited information, e.g. few datapoints or noisy data, by applying the general argument of maximum entropy, [94, 95] which corresponds to basic thermodynamic principles of the second and third law of thermodynamics. This can be applied to the general field of information reconstruction and interpretation in a general manner. [96] In the crystallographic context, this is applied by calculating the electron density distribution based on the known reflections, e.g. low-resolution datasets of proteins, which fulfills this maximum entropy criterion. Several implementations are reported, e.g. *BayMem* [97] or *ENIGMA*, [98] and are capable of refining proteins to a good extent. [99] But since these methods are very costly in terms of computational resources they were not considered in this thesis and will therefore not be described in further detail.

1.1.9 Wavefunction Based Refinements

Acknowledging the huge success of QM calculations of wavefunctions (for theoretical background see section 2.2) in describing features of molecules ranging from bonding, inter-molecular interactions to absorption spectra etc. and the advances in the prediction of molecular, but also solid-state structures, and their corresponding properties, it would be desirable to use these advantages in the refinement of X-ray diffraction data. The biggest advantage of wavefunction calculations is the fact that they can yield results *ab initio*, that means there is no input required except for some pre-defined basis sets to model the atomic orbitals. Given this information, the calculation, by definition, will yield a wavefunction that describes the molecule according to the best knowledge of quantum mechanics. This way the density that can be calculated from a wavefunction can be used and interpreted to model a crystal structure taking into account the density distribution arising from the calculated wavefunction. This way, the only user-selected input is the level of theory used. By combination of the fields of QM and Crystallography to perform more advanced analysis and models of the crystalline materials, the field of Quantum Crystallography emerged (see section 1.2)

The *Hohenberg-Kohn* theorem states that there is only one wavefunction that corresponds to the ground state density of a molecule. This is expressed by the fact, that the potential of the molecule is uniquely defined by its charge density. [100] If the potential is a unique property of the density, the many-body wavefunction must follow this unique relationship. The second aspect of the theorem provides the insight that if the density used for the construction of the wavefunction is the true ground state density the energy will be the absolute minimum of the wavefunction. [101] By application of the Hohenberg-Kohn theorem, it should be possible to find a wavefunction which corresponds to the ED distribution measured by X-ray diffraction, if this measured density is the ground state density.

1.2 Quantum Crystallography - A Brief Definition

As *Debye* stated in 1915, "the experimental study of the scattering of atoms, in particular for light atoms, should get more attention, since along this way it should be possible to determine the arrangement of the electrons in the atom." [12] The field of charge density determination and analysis tackled this problem and was successful in deriving a variety of descriptions for the distribution of electrons in the atom and bringing the gained knowledge into chemical and physical context. [74] The analysis of topologies of electron density, derived electrostatic potentials and even spin densities of unpaired electrons linked the high-quality diffraction experiment with the understanding of chemistry and molecules. Following the line of thought of the charge or electron density analysis, the field of Quantum Crystallography arose. Although there is a debate in the community about the exact definition of the field Quantum Crystallography (QCr), a small summary of the purpose of the field is necessary to understand the different approaches it provides and how they can be improved and applied in this thesis.

The original pioneering ideas of Quantum Crystallography stem from *Massa* and *Clinton* based on their work on the direct determination of pure-state density matrices published as a series of five publications. [102–106] In the following decades this work was extended and applied to constrain calculations by accounting for measured X-ray scattering data to obtain one-electron density matrices. [107–111] The development of nearly linear scaling methods, the kernel energy methods (KEM), was the result of these efforts. KEM successfully obtains density matrices and properties for systems like DNA, RNA and proteins. [112–114] In this context, the term Quantum

Crystallography was coined by *Huang, Massa* and *Karle*. [115, 116] In 2017, *Grabowsky, Genoni* and *Bürgi* summarized the definition of this term by highlighting two aspects: [117]

1. Enhancing the insight into the crystallographic structural properties and accuracy of structural models employing QM methods.
2. Enhancing the results of QM methods using the information present in experimentally measured structure factors.

The topic of quantum crystallography has drawn a lot of attention in the literature and shows a broad variety of possible applications. [116, 118–123] The two aspects of quantum crystallography are different approaches to make use of the knowledge of two mature fields of chemistry: quantum chemical calculations and crystallography, to improve each other.

Application of the Aspects of Quantum Crystallography in Different Methods

In the first aspect mentioned above, the knowledge about chemical bonding, regardless what type of bond, is used to improve the model of a molecule used for the explanation of the observed diffraction pattern. This information about bonding and atomic interactions and their effect on the electronic structure of the atom is available for example by QM methods and can be used to improve the basis of the model. Since the refinement of a structure is always based on a model for the electron density of the molecule, the accuracy and precision of results are influenced directly by the assumptions made in the model. If a better model is used the refined structure becomes more accurate. The main reason why this type of information was not used previously is the lack of computational power and general applicability of the corresponding models at the time the models for spherical refinements were established. [59]

The second aspect of quantum crystallography uses experimentally available information about the distribution of electron density to model a wavefunction which incorporates these measured features that *ab initio* methods fail to describe. [122] This has become possible with the advances of modern diffractometers and high-quality diffraction experiments available even at conventional lab-sources. Examples of effects that are not correctly modeled by *ab initio* calculations is the unknown exact exchange correlation functional in Density Functional Theory (DFT) (see section 2.2.3.4) or the description of mixing of excited states with non-ground state systems as it might very possibly be present as an average over a complete crystal under investigation by diffraction experiments. Another approach might aim to obtain more accurate methods usable for bonding analysis, prediction of reaction pathways or, in the case of this thesis, the modelling of protein–ligand interactions.

Quantum Crystallography in different refinement techniques

In principle, even the IAM can be regarded as an early attempt of a quantum crystallographic method since the tabulated scattering factors are obtained from QM calculations of isolated atoms even including relativistic corrections. [124] Despite the crude spherical approximation, these atomic wavefunctions allow structural refinements, in principle fulfilling the criteria for aspect 1 of quantum crystallography. [41] The same way one could consider the Multipole Model as a quantum crystallographic method, but only with respect to the first aspect because although the ED distribution may be refined, there is no direct insight into properties of a wavefunction or enhancement of wavefunction properties. Concerning multipolar model database techniques the scattering factors for the multipoles and the core are obtained from pre-calculated wavefunctions or tables for isolated atoms.

Techniques like Hirshfeld Atom Refinement (HAR) [125, 126] (see section 2.1.8), where a QM wavefunction is used to calculate tailor-made atomic scattering factors, fulfil the first aspect of Quantum Crystallography, while the usually subsequently executed X-ray Constrained Wavefunction fitting (XCW) [127–129] (see section 2.1.9) is a method of the second aspect, since it calculates a wavefunction minimizing the difference to observed diffraction patterns while minimizing the wavefunction energy. This makes X-ray Wavefunction Refinement, which is the combination of both techniques [92, 130] (see section 2.1.10), a method fulfilling both aspects.

1.2.1 Hirshfeld Atom Refinement, X-ray Constrained Wavefunction Fitting and X-ray Wavefunction Refinement

In these methods developed by *Jayatilaka et al.*, the wavefunction of a molecular fragment of the crystal is calculated and used to partition this molecular electron density based on the Hirshfeld stockholder partitioning scheme [131] (for details see section 2.1.8) into atomic contributions. These atomic densities are Fourier transformed with respect to the crystal lattice to yield the atomic form or scattering factor f. [125, 127, 131] These are used instead of the tabulated scattering factors in the IAM or the multipole populations in the Multipole Model to calculate the structure factors for the least-squares refinement and subsequent analysis.

This approach is used in two different ways, that correspond to the two aspects of QCr, as defined above. The first way is the use of improved scattering factors in HAR to refine position and displacement parameters and obtain the best possible structural description of the diffraction pattern. In principle, the wavefunction used to obtain these scattering factors can be pre-calculated or provided by different well-established QM-software, as long as the partitioning and refinement engine can read them. Applications of this interface between different wavefunction generators and the refinement engine are the combinations of Extremely Localized Molecular Orbitals (ELMOs) [132, 133] with HAR. [134]

The second procedure, XCW, [127] models the observed ED by modification of Molecular Orbital (MO) coefficients to match observed structure factors depending on their measured certainty. Scattering factor calculation of the electron density obtained from the molecular wavefunction convoluted with thermal motion parameters of atoms is used to compose calculated model structure factors, which are then compared to the experimental diffraction data. A combination of both, where first a HAR and a subsequent fitting of the wavefunction using XCW is called X-ray Wavefunction Refinement (XWR). [92, 130]

There is a subsequent iterative refinement of structure and wavefunction, where the resulting fitted wavefunction is used instead of a newly calculated *ab initio* wavefunction during HAR, to have a better description of the experimental ED. This method is called total-XWR. [135] But a simultaneous refinement of MO-coefficients and molecular structure is technically not possible in the framework of this refinement technique. It should be mentioned that there exists literature on the refinement of crystallographic data using periodic Bloch-type-wavefunctions in contrast to the molecular approach used in XWR. [136] However, this approach is not part of this thesis and will therefore not be explained in further detail.

A different approach to obtain wavefunctions from the diffraction data is the use of kernel energy methods (KEM) which divide the problem into smaller pieces that allow easier independant calculation and can also be used to calcualte densities and minimize the difference between model and observed electron densities. [111, 137, 138] While this approach is based on the direct projection of the observed electron densities into the density matrix of a wavefunction the methods by *Jayatilaka et al.* do not constraint, but rather restrain the wavefunction to electron densities similar to those observed by diffraction.

Advantages

The wavefunction based approach has advantages inherited from their respective field of origin. One is the separation of the thermal description of atomic motion from the calculation of the ED. Since a description of the model using the wavefunction derived densities inherently describes the bond density, usually neglected by the IAM, the ADPs will become less biased by the usually present residual density on bonds, trying to model bonding by displacement of the atoms, rather than separating these two effects *a priori*. This avoids the significant correlation of parameters in free multipole refinements of the elctron density with simulatious structural refinement.

Secondly, the additional side-conditions during the calculation of the ED from a wavefunction prohibits numerically possible but unphysical densities. This is ensured by the boundary conditions during QM calculations (compare section 2.2.1 and equation 2.39). The wavefunction needs to fulfil the criteria of common theoretical chemistry, for example for electrons to be present in space. The multipole technique allows the introduction of regions of negative density of one type of atom to match the measured density better if, for example, another atom provides too much density at that position. This behaviour could be understood as a cancellation of errors, where two physically wrong atomic densities end up describing the observed density. This behaviour is impossible using wavefunctions and Hirshfeld partitioning. This ensures physical behaviour, even if some assumptions are made. In summary this prevents overfitting of the model.

Lastly, the availability of all descriptors used in theoretical calculations, ranging from charges over chemical bonding analysis to individual MO population and analysis, is favourable. There is a wavefunction either being used during the process or being its result. This makes it easy and convenient to evaluate the bonding and interaction features of the structure under investigation without additional calculations. Additionally, the flexibility to include as much sophistication into the level of theory as one wishes the level of detail used in the constrcucted model is only limited by the quantum mechanical method used, computational resources and diffraction data quality. Since the latter has significantly improved compared to the available data when the IAM was introduced, [47, 139] it is about time to improve the models and their convenience to be used to treat this data.

1.2.2 Hydrogen Atoms in XWR

As a consequence of the accurate and physically correct non-spherical description of the ED, the determination of hydrogen atom positions and thermal parameters is not only possible but also very accurate and precise, since intrinsically the wavefunction locates the bonding electrons naturally. [140] To show the difference between the IAM and the tailor-made HAR scattering factors the ones of a carbon atom and a hydrogen atom in the epoxide molecule were plotted in Figures 1.6 and 1.7.

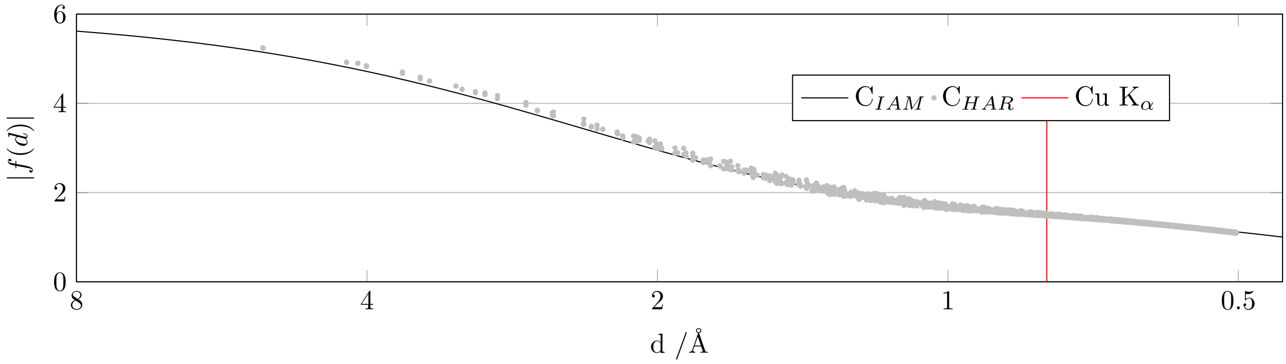


Figure 1.6: Plot of the value of the scattering factor with respect to d-spacing for a neutral carbon atom and a carbon atom in epoxide in HAR. [41] Red line representing the diffraction limit of copper K_α radiation plotted to visualize the importance of differences using one of the most common radiation radiation types for in-house measurements and beyond this limit.

The difference between the spherical model and the Hirshfeld atoms is dominated by the partial charge one would expect based on electronegativity. [141] It leads to a partial negative charge of carbon, yielding an increased low-resolution scattering factor, while the partially positively charged hydrogen atom has smaller low resolution scattering factors. At medium resolution around 1.5 - 1 Å, the difference to the spherical scattering factor has a wider spread, that can be explained by the difference of density depending on the direction of the scattering vector. This spread could be interpreted as a measure for the non-sphericity of the atomic electron density. For carbon, there are form factors that go well below the spherical scattering factor at the identical resolution. This means the electron density contribution to this specific form factor is lower than that of neutral carbon, despite the partially negative charge of this atom. On the other hand, there are individual form factors in the selected hydrogen atom of epoxide, that lie above the spherical, neutral scattering factor, even though one would expect this hydrogen atom to be partially positively charged, leading to even lower scattering contribution. This is due to the fact, that compared to the spherical hydrogen atom the wavefunction derived density is shifted into the bond between C and H. This is reflected by some scattering factors, that correspond to the directions of scattering caused by the density in the bond, to be higher than in a spherical atom. The red line, symbolizing

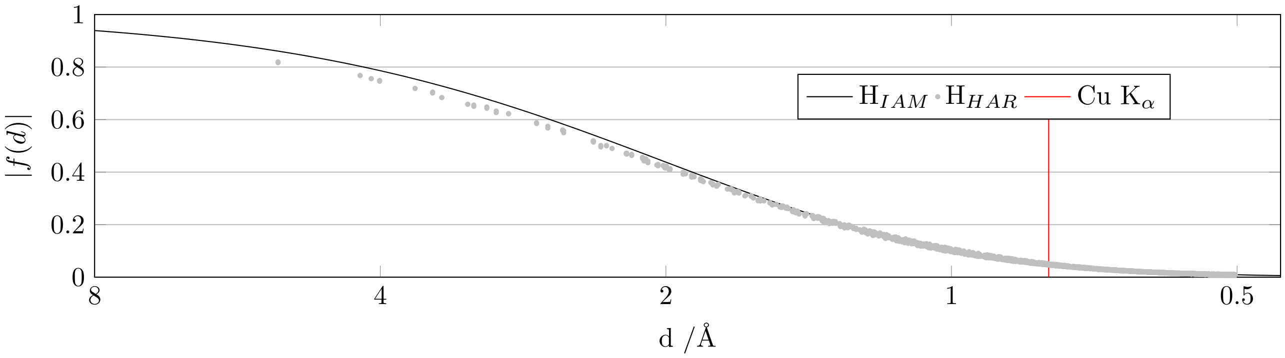


Figure 1.7: Plot of the value of the scattering factor with respect to d-spacing for a neutral hydrogen atoms and a hydrogen atom in epoxide in HAR. [41] Red line representing the diffraction limit of copper K$_\alpha$ radiation plotted to visualize the importance of differences using one of the most common radiation radiation types for in-house measurements and beyond this limit.

the common diffraction limit of Cu Kα radiation, which is very commonly used in single-crystal diffraction experiments, is shown to visualize, that the effect due to non-spherical behavior of atoms is very well observable, especially due to charge distributions, from even lower resolution diffraction experiments.

1.2.3 Restrictions in Wavefunction Based Refinements

Speed

The most severe drawback of XWR is the high computational cost associated with the repeated calculation of wavefunctions. Although these methods can be parallelised to a certain degree, the complexity rises much quicker with the number of atoms N for normal methods. [142] For example, proteins or some heavy elements surrounded by bigger organic ligands in metal-organic complexes might easily exceed the number of basis functions most computers can accommodate. While there have been advances in the field of linearly scaling DFT methods or AI-assisted density predictions (for example [142–151]), these methods either are not easily available or suitable for the general refinements of all types of molecules.

Here, the approach of separating the problem in a "divide-and-conquer" approach can help. Two approaches will be mentioned here: the kernel energy methods and the orbital localization in extremely localized molecular orbitals. [111,116,121,133,137] Both work by separating the problem of a wavefunction calculation into smaller, more easily solvable calculations, paying the price of losing general validity of the wavefunction in terms of orthogonality or accuracy. However, these smaller problems could be solved in a highly parallel manner, which even allows the Graphical Processing Unit (GPU) of graphic cards to be used to evaluate the smaller problems. This would be beneficial since their setup is optimized for the solution of high numbers of small, independent problems.

The ELMO procedure was implemented in conjunction with the framework of HAR and is called HAR-ELMO. [134] Some previous results obtained for proteins will be shown in this thesis later. The benefits for different classes of compounds and the practicability in general use of HAR-ELMO still needs to be elucidated, as the number of available pre-calculated ELMOs is currently very limited. This leads to the same type of problems as there is for Multipole database techniques. An approach of tailor-made QM-ELMOs, which are calculated on the fly, to be used during the refinement would solve this issue, but lose the great advantage of instantaneous availability of wavefunctions, that stored ELMOs would provide. [152]

A different approach to tackle issues of speed is the implementation of more optimized, modern software that makes the calculations more feasible using the existing methodological framework. The new methods presented in this work rely on modern QM-software in combination with approximations during the calculations that allow enormous speed-ups at almost no loss of accuracy. Additionally a new partitioning software that is implemented in a modern language like C++ in a highly optimized manner allows significant speed up while using higher accuracy for the calculations required.

Disorder

Although, XWR seems to provide superior results for different kinds of molecules, its application is limited. Disorder of the molecule of interest is conceptually not possible to integrate into the calculation of a single wavefunction. Partial occupation of atoms and the exclusion of repulsion arising from different geometries or parts of a disordered molecule are impossible to implement during the wavefunction calculation. A normal disorder model, as used in the refinement of crystal structures, places atoms in close vicinity of each other. In normal refinements this is not a problem, as the scattering factor will simply be multiplied with the occupation and no interaction between any atom is taken into account. In wavefunction based techniques this is not the case. The atoms placed in vicinity to each other will interact. This will lead to non-physical behaviour in the calculation of wavefunctions, if both atoms of the disorder model are present.

The only solution would be a calculation of individual molecules corresponding to the different disorder-parts to be modelled but the setup of calculations like these is not possible in *Tonto*, [153] which is the only software that can perform XWR. Also considersing partial occupations on otherwise not disordered atoms, for example in cavities of a crystal being filled by guest molecules, is impossible, for the very same reason. Disordered solvent molecules would be required to be excluded from the structure factors observed by algorithms like *SQUEEZE* or *BYPASS*. [32, 33] Although this technique is available in common IAM packages, an interface or implementation in wavefunction based refinement software is missing, so far. In principle, careful manual application removing the solvent contribution to the diffraction pattern is possible but neither convenient nor time-efficient. This would also not correspond to the correct physical model of the situation but a deconvolution of the disordered area from the ordered part of the structure. It is just the last resort if all other model procedures either fail or are unavailable. Additionally, these approaches do not allow for a re-evaluation of the solvent contribution after the refinement, which might be required due to the change of the size of a molecule, if the hydrogen atoms become better described at bigger distances compared to IAM refinements.

Heavy Elements

To this date, it is necessary, that the applied basis set for the calculation during XWR is an all-electron basis set since effective core potentials would lead to a lack of core electron density which is crucial for the calculation of scattering factors. Also until recently, the calculation of wavefunctions of heavier elements was complicated or almost impossible, due to the lack of proper basis sets. A recent development allows the use of all-electron basis sets of the x2c-family for elements up to Rn. [154] Additionally the jorge-family of basis sets [155, 156] in both relativistic and non-relativistic formulation were made available for XWR in the course of this work. With these basis sets on different levels of sophistication, that is single zeta, triple zeta and polarized triple zeta in the case of x2c-family and double- and triple-zeta both in relativistic contracted and non-relativistic version for jorge-basis sets, the calculation of all-electron wavefunctions became possible for the purpose of XWR.

But the measurement of crystal structures of heavy elements is complicated in terms of the experimental conditions. The heavy elements have very high absorption coefficients for X-rays and usually compounds containing heavy elements decompose or lose crystallinity during the experiment. This effect is called radiation damage. It is quite well documented for protein crystallography, but the intrinsic mechanisms and treatment for small molecule and especially heavy element compounds are not well described in the literature. [157–163]

High Symmetry and Network Compounds

Due to the refinement algorithms inside *Tonto*, [153] which is the current software for XWR, the refinement of atoms residing on special positions in the unit cell is unstable. A special position is a coordinate in the unit cell, where symmetry constrains the possible displacement and position of atoms. If, for example, an atom is positioned on an axis along the c-direction of the unit cell that coincides with a two-fold rotation axis the atom position can vary in the direction of c, but moving it off the axis in b or a direction would lead to a symmetry-generated atom on the opposite side of the axis. The only way to move the atom of that axis is to solve the structure using a lower symmetry space group, which does not have this symmetry element, if the displacement of the atom is required.

In *Tonto* the refinement of atom positions and displacement parameters is attempted by applying the site symmetry of an atom onto the shift vectors which are basically the change of parameters required to obtain a better fit in the least-squares procedure (compare section 2.1.5). This should in principle yield a stable behaviour of atoms on special positions, if for example a movement off the axis was generated by the least-squares procedure, would cancel out after applying the site symmetry to the shifts. Unfortunately, this does not always work in *Tonto*, which is why sometimes displacement parameters are refined, that are forbidden by symmetry or atoms leave the special position, leading to an unstable refinement, as in the next refinement step the contribution of two atoms of the same type will make the refinement break. These issues are the same as in IAM refinements, but the algorithms in common refinement engines have special treatment for these cases, where the values are not refined. [50, 164–166] This is solely a limitation of *Tonto*.

A different restriction, arising usually also for highly symmetric structures, is the fact that a network compound, that is a compound where it is not possible to define a unique molecular entity which can be differentiated from its neighbors is practically impossible to refine. This is due to the fact, that one has to calculate a wavefunction which is performed in *Tonto* in a normal molecular context. This means that in a network compound, where no molecular boundary is defined, bonds have to be cut in order to calculate a wavefunction. In the case of ionic salts or coordination compounds, it might be possible to calculate the wavefunction for a big assembly of many asymmetric units and only evaluate the wavefunction in the asymmetric unit.

This is not possible in the case of organic periodic networks compounds like polymers. Here the only solution would be the implementation of periodic boundary calculations, where the wavefunction is calculated for the whole unit cell or the asymmetric unit and the neighbouring atoms introduced by the use of atomic orbital functions or plane waves, which have the constraint of steadiness at the periodic boundary. These calculations are more complicated in analysis and significantly more time consuming, but could not only solve the problem of covalent network compounds but also lead to correct modelling of the crystal field effect. [167–171]

1.3 Complementary Bonding Analysis

Complementary Bonding Analysis, a term my PhD colleague Dr. Malte Fugel coined, [172–175] is the combined use of various descriptors for chemical bonding to allow a more generalized understanding of the nature of the interaction between chemical entities. This is necessary due to the absence of a direct "bonding-operator", that could define the chemical bond uniquely. A variety of methods for chemical bonding analysis have been developed, as summarized in Dr. *Fugel*'s work. [172–176] In this thesis, the attention will be drawn onto the descriptors used within the framework of the reported work. This includes different methodological approaches, which will be shortly introduced in the following sections:

- Real Space Descriptors (e.g. electron density, Laplacian, ELI-D, NCI and their topological analysis)
- Orbital Space Analysis (e.g. NBO)
- Energy Space Analysis (e.g. EDA)

The conversion of files to allow these very different kinds of analysis without repetition of QM calculations in different software is a big issue and motivated the development of the software presented in this thesis. In order to highlight the kind of description and properties these descriptors can show, in the follwing paragraphs a quick summary is given. For more details please refer to section 2.3.

1.3.1 Real Space Descriptors

Real-space descriptors include the scalar and vector fields of the electron density, its gradient and Laplacian as well as the localization or localizability of the electrons. The electron density and its derivative is used in the framework of the Quantum Theory of Atoms in Molecules to define atomic basins. (compare also section 2.3.1 and [177–180]) These can be analysed in terms of their properties, including bonding indicators like the density and curvature of the density at the so-called bond critical point or the charges of the atoms and the contribution of density of one atom to a different one as expressed by for example the Delocalization Indices or Source functions. [179, 181, 182]

Additionally, it is possible to look at the reduced density gradient in order to get insight into especially non-covalent interactions (compare section 2.3.3 and [183, 184]).

A different approach looks at localization of electrons, namely the Electron Localization Function (ELF) and the Electron Localizability Indicator (ELI). [185–187] The concept of localization can be understood as the degree of isolation of the electron density in a certain region. This property has a high value in regions which coincide with valence chemical features like lone pairs or chemical bonds, which allows conclusions about bonding features.

Volumetric Data - From Atom Shells to Proteins

The real space descriptors of chemical bonding mentioned above are often evaluated in volumetric datasets. These are voxels, in most cases spread in a uniform manner, which are assigned a certain value at each point in space. The definition of these 3D grids is often appealing to computational chemists since their setup is not very much different from the calculation of a property at a certain point in space. The same algorithms only need to be repeated in a loop-wise manner over all voxels. This is easier than the need to find an analytical expression of a property to plot or evaluate, since these numerical evaluations also allow interpolation and easier visualization, as the calculation of isosurfaces and maps can be performed quickly and from the existing data without further evaluation of the wavefunction.

These points in space of the voxel-mesh can easily be created by defining four three dimensional vectors and three iterator sizes: an origin of the voxel space, unit vectors for the voxel space, all in reference to the Cartesian space, and the number of points in which each edge of the voxel space should be separated. But due to the nature of this approach, the number of calculations rises with N^3 in terms of the spatial resolution, which is highly impractical for big systems such as proteins or other polymers (see section 4.4). On the other hand, an extremely high resolution is sometimes necessary to understand the very detailed effect onto properties like the electron density or the Laplacian of the electron density around heavy elements to understand secondary interactions (see section 4.3).

Parallelisation

Since the task of calculating these voxel meshes is very independent for its evaluation at a given position, as no datapoint calculated has any effect onto any other datapoint, the calculation can be parallelised in an almost linear fashion with nearly no additional cost on memory resources. The input for these calculations, like wavefunction coefficients or multipole parameters, are unaffected by the calculation of properties, so they can be shared between individual calculations of all data points. Possible targets for code development include Graphical Processing Units (GPUs), as well as enormous computing clusters with hundreds of Central Processing Units (CPUs). Especially GPUs are desirable for these tasks since their very design is intended to tackle great numbers of relatively simple tasks, like the evaluation of gaussian type orbitals of wavefunctions, at the same time. Compared to CPUs the individual computation might take slightly longer, as the base clock speed and handled instruction complexity per base clock cycle on GPUs is lower, but the number of threads and the significantly lower price per thread of GPUs make it much more economically and time-efficient. [188]

Definition of the volume of interest

However, these calculations require the user to input the voxel space to be evaluated either by hand or by the definition of at least 3 atoms to be investigated individually. This selection defines a vector system which can be used for the definition of gridpoints. If one wishes to investigate several atoms or interatomic regions, the programs available would require the user to set up each calculation individually. An automatic procedure e.g. to sample over the whole molecule by automatically calculating the required portions of atoms or bonds was not available yet, but will be presented in this thesis in section 3.3.3.

1.3.2 Orbital Space Analysis

The picture of canonical molecular orbitals is difficult to interpret in terms of chemical reactivity or bonding strength. This is why different localization schemes for the transformation of the canonical orbitals were proposed. One of them includes the framework of Natural Bond Orbitals (NBO), [189–191] which tries to find that linear combination of MOs, which yields a bonding picture as similar as possible to that proposed by *Lewis* [192, 193]. This way orbitals which resemble simply interpretable chemical features are obtained. They coincide with classical bonding pictures like 2-center σ-bonds, localized π-bonds or metal-ligand orbital simplifications like σ or π back-bonding. Non-Lewis delocalization phenomena become available as perturbation terms in this framework.

1.3.3 Energy Space Analysis

The analysis of the strength of the interaction between two molecular entities in terms of energy space allows the estimation of energy associated with a certain bonding situation. There are different schemes which try to achieve this, wherein this thesis the *Ziegler-Rauk* approach was used in the form of Energy Decomposition Analysis (EDA). [194, 195] The calculation of infinitely separated fragments of a molecule and the combined complete molecule, which makes three required wavefunction calculations, is used to yield the total interaction energy as the difference of the individual fragments to the total energy of the complete molecule. Furthermore, the individual contributions in terms of individual contributions of Pauli repulsion energy, electrostatic energy and orbital interaction energies allow more distinct identification of individual character of a bond, especially when compared in a series of compounds.

1.4 Applying Quantum Crystallography in Drug Design

The principles and techniques mentioned above lay the foundation of manifold tools to be used in chemical or biochemical applications. In the following, the topic of drug design will be introduced, which is one of the major fields of biochemical interest.

1.4.1 Protein-Ligand Interaction

In order to rationally design drugs, it is crucial to understand the mechanism how a drug achieves its effect. This mechanism is called the mode of action. [196–198] In most cases, a drug either binds to a receptor or protein in an active site and activates or inhibits the effect of this protein or connected apparatus. For this to happen, the drug needs to be compatible with the active site in various ways, involving steric bulkiness and electrostatic complementarity. [199–204] The ability of a drug molecule to actually reach the tissue in the body where the target protein is present is also a significant factor to be considered during the development of a drug. For example, the brain is not reachable by all drugs due to the brain-blood-barrier. [205, 206]

In the case of irreversible interaction with a protein, the interaction is often achieved by binding the ligand through a covalent bond, which is formed when the molecule is at the right position in the active site. Other irreversibly bound drugs achieve the irreversibility of their binding through extremely high affinity and better fitting interactions with their target enzyme than the original substrate of the protein for this protein has or by irreversibly destroying the functionality of a protein, for example through a change of conformation. This kind of inhibition is much stronger than reversible inhibition where the bound drug is in equilibrium with other substrates or drugs, but might also have more side effects. [207] Irreversible inhibition makes it impossible for the protein to fulfil its original purpose in most cases. This will make the cell deconstruct the protein and needs a new protein to be synthesized for the original functionality to be regained. This might be beneficial in case of cancer treatments or rather invasive treatments but in cases like pain relieve or short-timed symptoms this kind of inhibition might be too drastic, so the strongest kind of interaction does not necessarily correspond to the desired effect.

1.4.1.1 Key-Lock Principle

One of the first models describing the mechanism of a drug targeting a protein was the "key-lock" principle, described by *Fischer*. [208] His observation of limited compatibility of yeast with different types of sugars opened the field of modelling protein-ligand interactions. A schematic representation is shown in Figure 1.8. This assumption in the model principally reduces the determination of interaction energies between a drug and the enzyme to a constant, simple expression. The energy associated with the docking could be approximated as the difference between the energies of both entities on their own and the energy in their bonded state. This would make the calculations of docking energies a simple task only relying on the calculation of energies of these three entities (drug, enzyme and drug-enzyme complex) while assuming a fixed conformation. Only those drugs that match the shape of the active site will show a negative interaction energy and therefore bind to the enzyme.

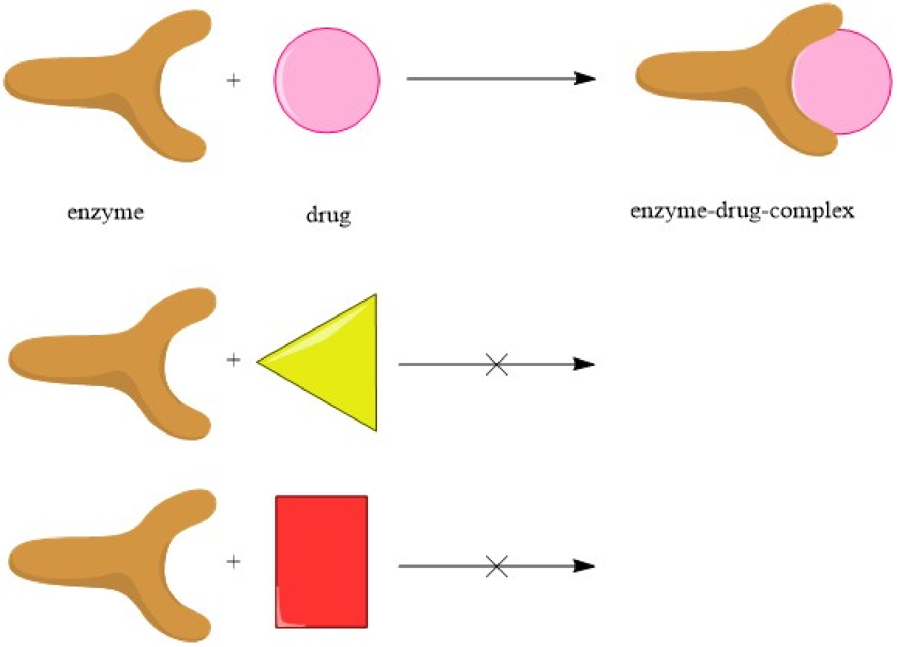


Figure 1.8: Schematic representation of the Key-Lock principle proposed by *Fischer*. [208] The enzyme has a pocket which only suits round targets to fit into the pocket. Targets of other shapes cannot bind into the enzyme pocket.

1.4.1.2 Induced-fit Mechanism

Fischer's concept was later refined and extended to take the flexibility of proteins and their active sites into account to formulate the "induced-fit" mechanism. [209] This mechanism describes additional factors like electrostatic complementarity and the possibility of proteins, upon the approach of a possible ligand, to change the conformation and form an open active site, which was previously not exposed. Additional factors leading to conformational changes of proteins include pH value, salt concentration, solvent composition and temperature. However, this process requires a certain amount of energy and therefore only the right kind of interaction upon the approach of a certain molecule will trigger these conformational changes. This makes the calculation of the interaction between a protein and a molecule more difficult, as the energy which was necessary to change the conformation of the protein into its active state is withdrawn from a possible gain in energy by binding a ligand.

1.4.1.3 Modern Protein-Ligand Analysis

Modern analysis of a known protein-ligand interaction is based on very different approaches. It is possible to perform docking simulations based on force fields (see below), experimental calorimetric measurements of a drug in a solution with the protein [210] or competitive experiments, where two drugs, one well known and an unknown drug, act in competitive inhibition and the relative ratio of binding is used to assess the relative binding strength. In the last case, the ratio of interaction is determined between the two competing drugs left in solution or bound to the protein by different analytic methods, for example, spectroscopy, mass spectrometry or NMR. [211–213] Also, the 50%-inhibition concentration, the so-called IC_{50}-value, may be used as a tool to assess the strength of a particular protein-ligand interaction. This drug-protein characteristic value is the concentration at which the activity of the target protein is only half of what it is without any presence of inhibitor. [214, 215] This can be indirectly interpreted in terms of binding strength, as a lower 50%-inhibition concentration corresponds to a more selective and stronger interaction.

Experimental methods have the significant drawback of a necessary synthesis and purification of the protein. This is sometimes very costly and complicated, while docking simulations only have limited accuracy or extent, due to inaccurate force field assumptions, missing parameters or due to restrictions in feasible time scales, as some simulations can take a rather long or require very sophisticated computational resources. Additionally, docking simulations can not predict the covalent inhibition of a drug, as the breaking and formation of bonds can not be described by classical force fields. This would require QM/MM-MD or complete *ab initio*-MD (see following sections for explanation), which is too resource-demanding in many cases.

1.4.2 Rational Drug Design

For decades, the design of a drug in an *a priori* manner has been the ultimate goal of many disciplines of chemistry. New drugs are always required for the treatment of new diseases or to minimize side effects of already established drugs during treatment. Some drugs are used in preventive measures to avoid the appearance or milden the course of a disease. As long as there is life there will be diseases. [216–219] The number of known and approved drugs is growing every year. [220] There are various approaches to find a new drug.

One of them is the so-called High-Throughput-Screening (HTS). [221, 222] This approach is based on the synthesis and testing of many substances by modification of known substances and following trends during the investigations to obtain the best possible candidate for a drug. This approach is often understood as the experimental approach to drug design. Its major drawback can be summarized by the analogy of the needle in a haystack, as described by *Gane* and *Dean*. [223] And by increasing the number of substances that modern ultra-high throughput methods can produce the hope is to introduce more needles in the ever-growing haystack.

On the other side, there is the approach of rational drug design, which uses computational or structural knowledge to make predictions which molecules are of interest. [223–225] *Gane* and *Dean* call this the antithesis to HTS approaches. With the growing power and spread of artificial intelligence in the scientific community, the rational part of drug design might completely shift towards the computer. [226] An introduction to selected methods used for rational drug design will be given in the following sections.

1.4.2.1 Using Theoretical Models

Structure of Molecules

The structure and composition of a molecule has been the tool for chemists to predict reactivity, physical properties and possible applications. One of the most successful models, still being taught to chemistry students and even in schools, is the Lewis-formalism to describe the structure and bonding in molecules. [192] To this day it is used in chemistry, extended and adapted to the current state of knowledge about, for example, dative bonding, but still following the basic concept of its first implementation. The link that can be drawn in these formulas between functional groups and chemical as well as physical properties is still valid and used to this day. A variety of approaches is still being developed to find better and more accurate correlations between structures, graphs or other descriptions of molecules and their properties. [51–53] The prediction of the structure of an unknown molecule in a given environment is, therefore, a crucial step to be able to predict or estimate its properties.

Geometry Optimizations

One way to obtain the structure of a molecule is the minimization of the potential energy, called geometry optimization. This method is based on the estimation or calculation of energies and forces acting on the atoms of a molecule and subsequent minimization to find the equilibrium structure of the molecule under investigation. [227, 228] Since these calculations of forces, regardless of their formulation, and subsequent minimization of the energy by structural alteration are beyond practicability for an evaluation by humans the success of the technique was directly linked with the emerging calculating power of computers. Various approaches for the estimation of forces acting upon the atoms of a molecule exist. The two most common schools of thought (classical and quantum mechanical) will be explained in more detail in the following paragraphs.

Force Field approaches: Molecular Mechanics (MM)

Theoretical chemists have created force fields based on classical mechanics to simulate the behaviour of a novel drug. Such a calssical force field describes the chemical bonding in molecules, as well as their interaction with each other, by simple physical approximations. [229] For example, harmonic oscillators or approximated dispersive interactions by Lennard-Jones potentials are used to minimize the computational effort required to describe realistically sized systems which might consist of bigger multi-unit proteins with thousands of atoms efficiently. [230–232] Limitations in these approaches are the dependency of results on the level of approximation used and the restrictions of applicability of the classical force fields due to missing parameters. Many possible drug candidates might not be described in an existing force field. This is the case either due to their chemical constitution, e.g. the presence of atypical elements, which have not been parametrized yet, or the existing parameters are formulated in such a general manner that derived properties are too inaccurate to draw significant conclusions.

On the other hand, more precise results in this framework would become too costly, because sophisticated calculations involving quantum mechanics or hybrid methods are required, which rise the computational cost so much that the approximation of a force field might become inefficient in comparison to full *ab initio* methods. First attempts to perform simulations on any system regardless of chemical composition were done in the Universal Force Field *UFF*, [233] which covers all available elements of the periodic table. This force field performs rather poorly when it comes to more unusual bonding situations, for example sterically strained rings, conjugated bonds or non-σ or π type bonds, for example, 3-center-4-electron bonds, as well as shallow landscapes of the Potential Energy Surface (PES), where the flexibility of the force field is too small to find the experimentally observed conformation of a molecule. It is an example of very generally defined parameters, which allow wide applicability paying the price of accuracy in more specialized situations. Very promising general anharmonic force field approaches include QM derived force fields like *QMDFF*, [234] which are on the downside significantly more costly in terms of timing and computational resources, especially for bigger molecules.

In this work the CHARMM type force fields were used for geometry optimizations of proteins. [235–237] It is a type of force field that has many types of common residues pre-defined and allows easy parametrization of missing components. A more detailed explanation is given in section 5.2.6.

Ab initio methods

Since the applicability of force fields is limited for novel compounds and the accuracy of results sometimes is too vague to make an unambiguous decision of possible candidates for a decision to invest into the synthesis and screening, the methods of *ab initio* quantum mechanics are used. [238–240] This field of research and methods is called Quantum Biochemistry. [241, 242] These methods can predict various properties of a potential drug, for

example pK_a, optical properties, aromaticity as well as potentials. [243, 244] One of these methods is the use of local density matrices and localization-delocalization matrices based on a combination of topological analysis and molecular graph theory. [245–247]
The use of electrostatic potentials, HOMO-LUMO energy differences and also full chemical characterization and bonding analysis available by quantum mechanical methods give much more detailed, accurate and precise result compared to force field-based methods. The major drawback is a significantly higher computational power required to achieve these results. Since modern computers become more powerful and the same computational power keeps becoming less expensive the possible applications of quantum biochemistry become feasible for bigger and bigger systems day by day.

Molecular Dynamics (MD)

The structure of a molecule in its equilibrium geometry already gives rise to many properties solely through the nature of the molecule itself. Examples of these properties are dipole moments, electrostatic potentials or the steric demand of space. Many processes, especially in the context of drug design, are not only dependent on the equilibrium structure, but are also influenced by thermal motion, pressure or the presence of solvents or neighbouring molecules. These processes can not be reflected by a minimization of the forces alone. The factor of evolution through time would be neglected.
This problem is overcome by the introduction of velocities and time into the calculations and a subsequent iterative solution of the equations of movement, called Molecular Dynamics (MD) (for detailed theoretical explanation see Section 2.4). This way, the atoms are not in their equilibrium position, but move through space and time of a simulated system. This description of motion allows simulation and increase in understanding of processes like protein folding, [248–250] molecular recognition of a drug or enzyme target [251–253] or the change of the molecular arrangement in structures. [254–256] Since some years the application of MD is also possible for complete viruses or cells thanks to large scale computational facilities. [257] These calculations combine several thousand computational nodes and hundred-thousands of CPUs to simulate some hundred millions of atoms to understand processes of diseases like Human immunodeficiency virus infection and acquired immune deficiency syndrome (HIV/AIDS) or Coronavirus Disease 2019 (COVID-19). [216–219, 258]

Molecular Dynamics using Novel Compounds

As mentioned above, the definition of a force field for novel compounds and unusual bonding situations is challenging and usually requires the development of a new set of parameters. In the context of this thesis, the attention is drawn towards organosilicon compounds (see Section 1.7). This inorganic bonding situation is widely unknown to the biological field of application of MD, and therefore no set of force field parameters can be found in the literature. Most force field parameters found in literature involve silicon- and silicate-surfaces. [259–261] The listed literature only gives a small selection of existing force fields for inorganic silicon parameters. But the situation of silicon bonded to carbon and especially hydrogen atoms is significantly different compared to the bonding situations involving silicates or pure silicon. In the scope of this thesis, the need arose to investigate further into the nature of such bonding situations and the parametrization of such force fields to allow simulations of the interaction between well documented and parametrized force fields of proteins and new organosilicon compounds.

1.4.2.2 Using Crystallography

Protein Crystallography

The deep understanding of the three-dimensional structure of proteins and enzymes that biologists and biochemists have are strongly based on the determination of the crystal structure of proteins by X-ray diffraction experiments. The field of protein crystallography has a crucial contribution to the modern ways of drug design. [262–265] The way modern protein crystallorgaphy is performed strongly relies on the advances made in experimental setups to produce the diffraction pattern and software to subsequently analyse it in the late 20th century. [266–268] These advances and wide availability of the three-dimensional arrangement of amino acids in proteins made it possible to identify active sites and common combinations of amino acids found in active sites of a specific class of enzymes. They can be found by analysis of complexes of the proteins with a ligand to identify, for example, active triads or binding sites and allowed the identification of potential targets through screening methods. [269–272] Other techniques to identify active sites include active site mutagenesis, that require a large series of artificially modified versions of a protein and the measurement of their activity. [273]

Chemical Crystallography

Another approach for rational drug design is the use of known structural motifs and chemical features of a substance to identify possible target proteins for inhibition. The known modes of action of drugs allow predictions of a new protein-molecule pair for inhibition by comparing the pockets of the target enzyme with molecular structures available from crystallographic methods. The close contacts, like hydrogen bonding, are taken into account and are compared to already observed arrangements in active sites of other drugs. In 1994, *Klebe* [274] showed a correlation between the conformation of a drug in the active site of its target and the structure in the crystal of the pure substance, which is the basic idea of structural recognition. This "nearly perfect correlation" was later independently confirmed by *Pascard*. [275] Their work shows how the statistical occurrence of certain conformations and atom arrangements in crystals of organic molecules resembles the distribution found in active sites of proteins. [276]

Additional predictions can be made by comparing the structural motifs in the drug itself or by comparing the nearest surrounding atoms with databases of proteins or drugs. The complementarity in terms of, for example, geometric arrangement of hydrogen bonding donor and acceptors of the protein and a drug are evaluated in these approaches and used to find matches. This way potential drugs can be identified by fully automatic comparison of the databases of known protein structures with a small molecule structure. In these approaches, the interaction energies between molecules and their target protein can also be calculated by applying schemes like the intermolecular perturbation theory. [277, 278] One of these schemes is used in *IsoStar*, [3, 279] where structural features and functional groups are recognized and both the Cambridge Structural Database (CSD) and the PDB are searched for reported structures in which these structural motifs are found. They are compared in terms of neighbouring atoms around this motif or the electron density effects shown in the interaction density calculated for this functional group. A different approach is based on the comparison of energies and shape matches based on promolecular or hirshfeld density. [280] These approaches can give first hints towards a possible target for further testing, but it is not possible to predict the actual interaction energy for an unknown substance or whether a certain molecule still fits the enzyme pocket with all structural features present or will even be recognized by the target enzyme.

Non-spherical Models

In their publciations, *Hibbs et al.* showed, that the modern methods for experimental charge density refinements were capable of showing electrostatic properties and valence shell features of biologically active molecules. [281] The work of *Luger et al.* showed a possible applications of modern X-ray ED methods for rational drug development and understanding of certain modes of action and chemical selectivity. By plotting the isosurfaces of the negative Laplacian of the density, colour coding the ESP onto ED-isosurfaces and by Hirshfeld surface analysis they could correctly predict the selectivity for attack by the active site of a protein onto the drug molecule. [282–284]

Many applications and investigations on the molecular properties derived by experimental determination of the electron density and the possibility to transfer this knwoledge are reported in the literature. [75, 76, 285–289] In a purely theoretial approach based on the total ED, looking at the search for so-called bioisosteres is performed by the identification of the biochemically relevant functional groups of a drug and a search for molecules that have similar or identical functional groups arranged in a similar structure. [193, 240, 290–292]

Quantum Crystallography

The tools that the field of quantum crystallography provides, including accurate evaluation of ED distribution and corresponding ESP in combination with correlations of properties of a drug with respect to intrinsic bonding situations and bonding descriptors (compare section 1.3), allow a more sophisticated prediction of a potential drug candidate. [119, 293] In this thesis, a new approach to combine the extended knowledge of a drug molecule, available from quantum crystallography, with the extensive predictive tools of theoretical simulation of the enzyme-drug complex will be presented and used to evaluate a possible bioisosteric drug-candidate for ibuprofen.

1.5 Cysteine Proteases

Proteases are a class of proteins, that cleave other proteins at specific points between amino acids, based on the active centre of the protease. Cysteine proteases cut a target protein in the peptide bonded backbone where the sulfur atom of cysteine acts as a nucleophile. This often happens in a catalytic Cys-His-Asn/Gln triad, where His first accepts a proton from Cys, then the sulfur atom of Cys can approach the peptide carbonyl-carbon. This leaves the carbon-nitrogen bond without resonating C-O double bond and leads to the abstraction of the transferred proton from His onto the nitrogen of the peptide backbone, cleaving the bond. The N-terminus is now free and leaves the protease. The thioester which is formed is hydrolysed by water and the cysteine thiol function is regenerated, releasing the C-terminus of the cleaved fragment as a carboxylic acid. [294] In some cysteine proteases, the Asn/Gln residue acts as a catalyst for the hydrolysis of the thioester.
This class of enzymes was used in this thesis during the investigations on the similarity of densities in various environments as shown in section 5.1.

Medical application

Cysteine proteases are essential in the human body, as they are responsible for senescence, apoptosis, lysosomal activity, immune responses, prohormone processing and control of the extracellular matrix, because they can control protein content of their surrounding media. Since these proteases can be activated or regulated, the biosystem can control the presence of other proteins. [295] In some cases, viruses can inject the cell with one big protein chain, which is cut into functional pieces by proteases, which is one reason, why one wants to be able to control the activity of proteases. If the body is in a diseased state which involves a lack of natural protease inhibitors, the proteases will damage healthy cells and bones. In these cases, the additional administration of protease inhibitors can regulate the activity back to normal levels. Additionally, the inhibition of proteases is used in the treatment of HIV/AIDS and hepatitis C, as these viruses rely on the body cells to reproduce, which is prohibited in the case, that no protease is available to cut precursor proteins into their active form. There are also hints towards possible applications in cancer treatment. [296]

1.6 Ibuprofen

Ibuprofen is one of the most essential Non-steroidal anti-inflammatory drugs (NSAIDs) and pain relievers known to mankind, as pointed out in the World Health Organization (WHO) "Model List of Essential Medicines". [297] Only the most crucial drugs to sustain minimal medicinal service are listed there. It is produced in a kiloton scale and the mechanism behind its medicinal application, its metabolism and side effects are well-studied . [298, 299] It is present in almost every household and plays a role in the everyday life of many rheumatic patients or for relieve of period-related symptoms for women. Its synthesis is fairly simple and efficient and the side effects are limited.

Mechanism of action

Ibuprofen works through competitive, reversible inhibition of cyclooxygenases, especially Cyclooxygenase II (COX-II). COX-II is present in damaged tissue of animals and humans and oxidates arachidonic acid into prostaglandins. In damaged or already inflamed tissue, prostaglandins trigger the inflammatory response of the body. In the case of ibuprofen, the synthesis of prostaglandin E_2 is reduced and the inflammation is not propagated that much. Many of the side effects of ibuprofen include gastrointestinal problems like stomach irritations and rise of acidity in stomach fluids, as well as renal toxicity and in severe cases, when combined with alcohol or other kidney challenging drugs, renal failure is associated with the inhibition of Cyclooxygenase I (COX-I). [299] Since ibuprofen inhibits both COX enzymes in a non-selective way and the potency of ibuprofen on COX-I is bigger, it would be desirable to increase selectivity for COX-II. There are many drugs known to be more selective for COX-II, for example, naproxen, diclofenac, phenylbutazone or even pure salicylic acid, but these have similar or much worse side effects and are more expensive in production. [299–301] Therefore, a drug with similar structure and low side effects by being more selective towards COX-II would be desirable. [299] An additional limitation in the application of ibuprofen is its low solubility in water. Only 21 mg of the substance can be dissolved in a litre of water. [302]

1.7 Carbon-Silicon Switch

To achieve better solubility, lower melting points or higher selectivity for a certain protein, the modification of drugs yielding better and better derivatives was performed already for centuries. Acetylsalicylic acid was invented to reduce the strong appearance of gastrointestinal damages observed with pure salicylic acid. [299] The addition of protecting groups or smaller modifications of the structure to achieve better interactions with the protein by e.g. introducing an additional hydroxyl function, removing a small functional group or exchanging one for another functional group of a previously known drug, was part of the repertoire of pharmacologists. [303–305]

Isolobal exchange and *umpolung*

A more substantial approach is the exchange of atoms in the carbon "backbone" of a drug through silicon. By isolobal exchange, [306] it is possible to exchange a carbon atom in the structure of a known drug with a silicon atom, as they are in the same group of the periodic table of the elements and therefore have identical numbers of valence electrons and similarly shaped valence orbitals. Therefore, it can be expected that the structure of the drug will not change significantly, except for a slight increase in molecular volume, as the silicon atom has a bigger atomic radius.

However, the electronegativity of silicon is significantly different from that of carbon, resulting in new polarities in various bonds, such as the Si-C and the Si-H bond. Since the electronegativity of silicon is lower than that of hydrogen, the polarization of the bond is reversed, which is called the *umpolung*-principle. [141, 307] In trimethyl-silane, which would be isolobal to the isobutyl-group, the change in polarisation is illustrated in Figure 1.9. It is clearly visible how the strong negative ESP in the vicinity of the silane hydrogen atom differs from the situation in iso-butane, where a small positive region is visible. Also, the slight difference in the size of the isosurface shows the effect of this isolobal substitution on the molecule, as the iso-value is identical. Especially the area around the silane hydrogen atom, which is much more electron-rich compared to organic compounds, is bigger compared to its carbon counterpart.

Carbon-silicon switch

Using this approach, it is easily imaginable that this newly introduced ESP difference between a carbon based compound and its sila counterpart will lead to a different behaviour in protein pockets or aqueous solution, where polarisation and electrostatic effects play an important role. Anticipating interesting changes in interactions with proteins, various groups, namely *Tacke, Franz, Lazareva et al.*, carried out this so called carbon-silicon switch in different compounds. [308–312] In this thesis, the application of the carbon-silicon switch to the mechanism

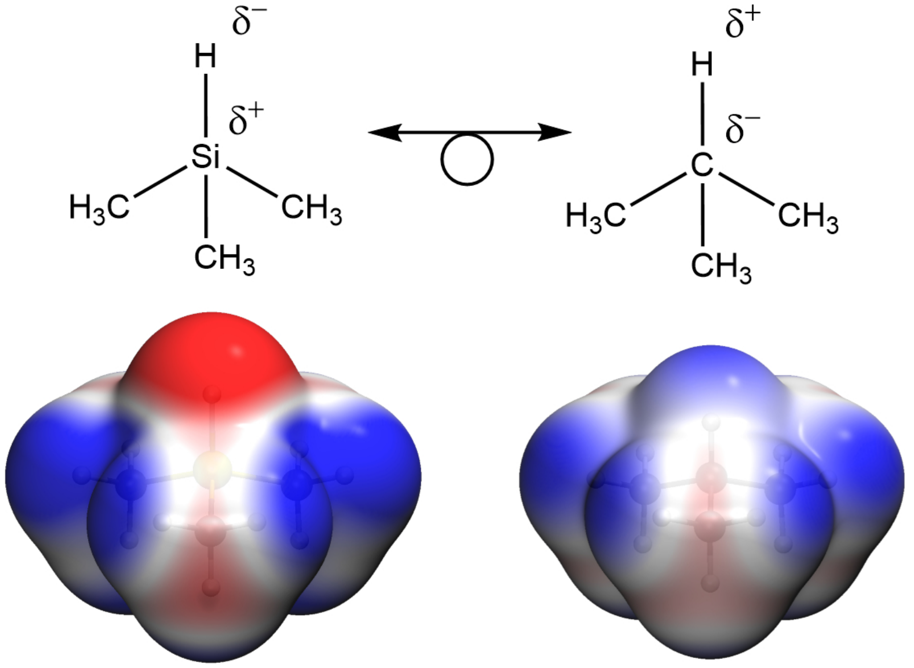


Figure 1.9: Visualisation of the *umpolung* in trimethylsilane (left) compared to trimethylmethane (right). Schematic Lewis structures with partial charges (top) and calculated electrostatic potential mapped onto the 0.001 a.u. electron density isosurface (bottom). Red areas correspond to negative potential of a positive test charge, blue areas to positive potential. Wavefunctions calculated with *ORCA*, volumetric data calculated using *cuQCrT* and visualized with *VMD*.

of action of ibuprofen by substituting the isobutyl group tertiary carbon atom with silicon is carried out and its effect onto chemical properties, protein-ligand interactions and selectivity for inhibition of either COX-I or COX-II is investigated.

1.8 Examples of Protein-Ligand Interactions in this Thesis

Usually, molecules are held in place in the active site of a protein by directed interactions, e.g. hydrogen bonds or covalent bonds. Both types of interactions will play a role in the course of this thesis. The drug which is modelled by the active site model compound (ASMC) as used in section 5.1 is called E64c. It binds to the active site of the cysteine proteases which it inhibits by forming a covalent bond between one of the carbon atoms in the epoxide rings and the sulfur atom of the cysteine residue, in the case of this work the active site of cathepsin B. [313–319] The epoxy ring is opened by the nucleophilic attack of the sulfur and the inhibitor covalently bound to the enzyme, as shown in Figure 1.10.

The enzyme can no longer perform its function, as the active site is now occupied by the former E64c molecule and the cystein residue, which is used in the enzymatic activity of cathepsin B to cleave the peptide bond through nucleophilic attack onto the carbonyl carbon atom of the backbone of proteins. [320, 321]

An example for the non-covalent binding of a drug to the target enzyme is ibuprofen, as used in section 5.2. It binds into the active sites of the COX target enzymes through the formation of two hydrogen bonds between the arginine residue present in the active site and the carboxylate group in ibuprofen. This complex is visualized in Figure 1.11.

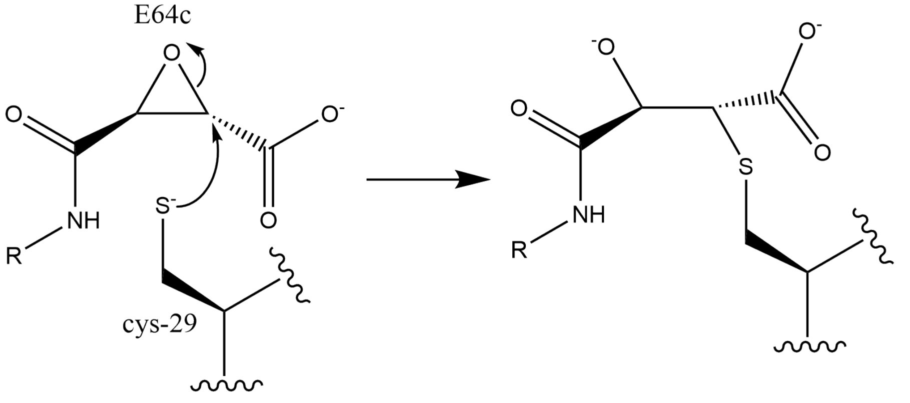


Figure 1.10: Scheme of the binding of E64c in the active site of cathepsin B by formation of a covalent bond between the cysteine and the drug molecule resulting in ring opening.

These two hydrogen bonds are strong enough to keep the molecule in place inside the active site, as will be shown in section 5.2, and occupies this bonding site for the approach of the usual molecule handled by COX, which is arachidonic acid. [299]

In the case of the normal substrates of a protein, the binding of this molecule in its binding site usually triggers nearby functional groups, the so called catalytic or reactive site to interact with the substrate and undergo a chemical reaction. [320] This can range from hydrolysis, over oxidation, methylation to metal addition. An inhibiting drug molecule often binds with the binding site of the protein, but is either incapable of undergoing the typical reaction or has steric constraints, that block the catalytic site. Other mechanisms involve adsorption of the molecule on the surface of a protein which undergoes conformational changes that make the active site of the protein inaccessible to the substrate due to the introduced structural change. Covalently bound drug molecules are often irreversible inhibitors, which means they remain in their position and render this protein molecule inactive permanently, while drugs bound by hydrogen bonds or other interactions usually inhibit reversibly, as this interaction has lower energetic barriers to remove the molecule again. These reversible inhibitors are often in competition with the substrate and are therefore called competitive inhibitors.

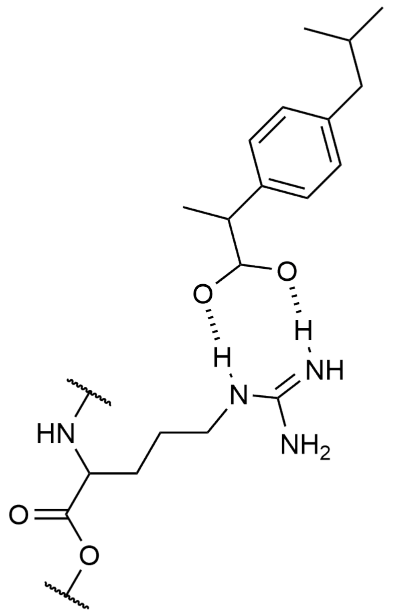


Figure 1.11: Lewis structure for the ibuprofen - arginine contact as the anchoring motive in COX-I and COX-II.

Chapter 2

Theoretical Background

2.1 Crystallography

2.1.1 Measurement

The measurement of the diffraction pattern of a crystal is a well established experimental setup, but requires precise machinery and sophisticated planning. The modern single-crystal diffractometer setup is usually controlled by computers and the rotation of crystals by fully automatic robotic goniometers. But to illustrate the required procedure to obtain a diffraction pattern the basic concepts and strategies will be summarized here referring to information from [20] and [21].

Due to the observation that X-rays diffract in the same way from crystals like visible light does from a lattice arrangement of slits, the Bragg equation [8–10] can be used to determine the distance between two lattice planes in the crystal. This is based on the assumption that the atoms in the crystal form different planes in three-dimensional space. The diffraction condition is met when Bragg's law is fulfilled:

$$n\lambda = 2d \sin\theta. \tag{2.1}$$

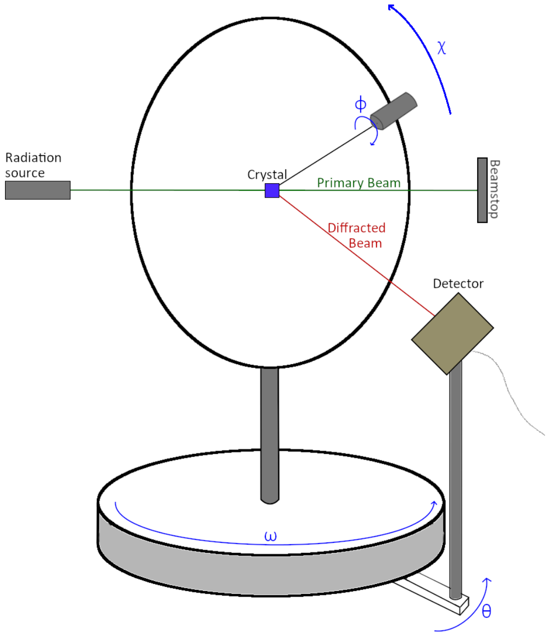


Figure 2.1: Scheme of the 4-circle apparatus to detect the diffraction pattern of a crystal. Parts labeled accordingly and the angles of rotation labeled on the corresponding pieces.

Here λ is the wavelength of the incident X-ray beam, n an integer number, d the distance between lattice planes, and θ the angle between the incident beam and the lattice plane. The lattice planes, that led to the diffraction are defined by the periodic arrangement of the atoms inside crystalline materials. The periodic arrangement defines the distance and angles between different Bragg peaks in the diffracted angle θ. Since the diffracted beam will leave the crystal with the same angle in opposite direction relative to the lattice plane the factor of two arises and the diffracted beam will have an angle of 2θ to the primary incident beam. To sample the whole 3-dimensional space of lattice planes it is required to arrange the crystal in all possible orientations with respect to the incident beam. This was chosen since it is technically easier to move the crystal in the beam than to move the X-ray source. This lead to the construction of the so-called 4-circle diffraction arrangement, where four axes of rotation are available to sample the whole space of 3-dimensional arrangements of the crystal and the detector with respect to the incident beam. A schematic representation of this construction is shown in Figure 2.1.

There are variations of this apparatus with, for example, a fixed angle of χ to prevent collisions between the goniometer apparatus and the radiation source. Careful rotation of the crystal using the three axes ω, χ, and ϕ is performed to sample all

Bragg peaks using the detector. The detector can be a point detector, for example an ionization chamber with a scintillation counter, that will evaluate the intensity of the diffracted beam based on the number of ionizations in the chamber. Modern diffractometers use area detectors that are capable of recording several peaks at the same time, which drastically shortens the time required to perform an experiment.
To plan an experiment and the best possible strategy to measure the diffraction pattern one needs to know which orientation the lattice has with respect to the lab-coordinate system, which is defined by the three angles of the goniometer and the distance between two Bragg peaks. This distance is inversely proportional to the size of the unit cell, the smallest building block of periodic crystals. This is usually measured on modern diffractometers by measuring test frames, searching for maxima on the recorded frame, and indexing them to find the three-dimensional lattice vectors between peaks. Knowing this expected distance between peaks and the orientation of the crystal it is possible to determine at which angles and detector settings the diffraction condition will be met and to optimize the measurement strategy.

2.1.2 Diffractometer

Modern diffractometers have made it possible to conduct high-quality experiments in a highly optimized construction, that, given sufficient crystal quality, can determine the structure of a compound within a few minutes. A labeled photo of the diffractometer used for some of the measurements is shown in Figure 2.2. Modern diffractometers are usually equipped with a cooling device, usually a nitrogen cryostream, that is focused at the position where the crystal is mounted. This is because the intensity of diffraction is directly influenced by the magnitude of atomic displacements, which can be lowered if the temperature is reduced (compare equation 2.5). Therefore a lower temperature yields a higher intensity of diffraction patterns. The cooling device restricts the angles that are possible for the goniometer to reach, especially in χ without provoking a collision with other parts of the arrangement. The arrangement with a circle that holds the mount for the ϕ rotation, as depicted in Figure 2.1, was dropped in favor of robotic arms, that restrict the movement in ω much less and allow more combinations of angles to be reached within the restricted available space around the crystal due to the positioning of the beam stop, radiation source and cryostream (compare Figure 2.2, labels 8, 1 and 3). If the difference in θ angles between two Bragg peaks is high enough an area detector can be moved very close to the crystal, allowing a broad simultaneous sampling of the diffraction space is possible. If the angular difference between peaks is small, which is usually the case for large unit cells, that is very long distances between the lattice planes (see also section 2.1.3), the distance between the detector and the crystal needs to be increased.

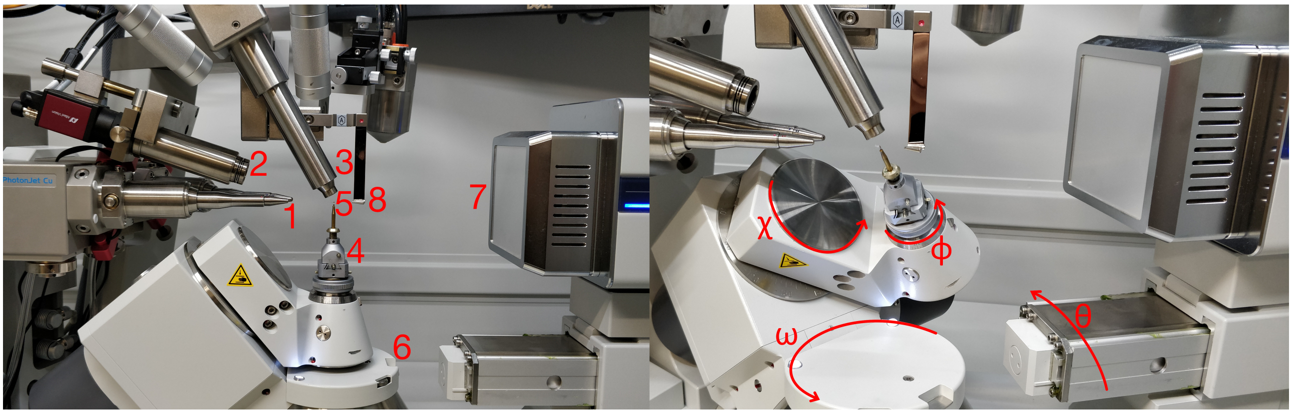


Figure 2.2: Photos of the diffractometer with numerical labels of the corresponding parts (left) and the axes used for rotation of the crystal and diffractometer during an experiment (right). The number is always on the right side of the corresponding piece of equipment, except for the detector. The label of the angle is always outside the circle of rotation. 1 = X-ray source; 2 = Camera for crystal alignment; 3 = Cryostream; 4 = Goniometer Head; 5 = Crystal Position; 6 = Goniometer; 7 = Detector; 8 = Beamstop.

The control of the robotic arms and detector readout is handled by a computer, that can calculate the optimal strategy to perform series of individual orientations and movements of the crystal to obtain all diffraction peaks expected from the found lattice parameters after initial screening in the shortest possible time.

2.1.3 Evaluation of Experimental Data

The diffraction data obtained by the X-ray detector needs to be processed to be of value for crystallographic refinements or analysis. Therefore it is necessary to perform a series of steps, that are usually done by software developed by the vendor of the detector. It transforms the photos or intensity signals from a detector into relative intensity signals, that are assigned an individual index of the expected Bragg-Peaks based on the orientation of the apparatus and corrects for systematic errors like background noise in the detection image or the peak profile, as the Bragg Peak, in reality, is not an infinitely sharp δ-peak but has a width, due to for example defects in the crystal lattice, the effect of limited crystal size and non-ideal coherence and focus of the incident X-ray beam.

Determination of Orientation Matrix and Unit Cell

There is no way to know exactly in which orientation atoms and molecules in a crystal are arranged *a priori*. By evaluating the diffraction pattern measured by the detector, one can determine the distance between measured intensity peaks and conclude the unit cell of the lattice defining the crystal periodicity. During this process it is also possible to determine the orientation of this unit cell with regard to the lab axis system of ϕ, χ, and ω, resulting in the orientation matrix. A unit cell consists of 6 characteristic metrics, that uniquely define classes of different lattice systems:

- α
- a
- β
- b
- γ
- c

a, b and c correspond to the length of the unit cell and α, β, and γ to the angle between these principal axes. In different combinations of parameters the following lattice systems occur:

triclinic	$a \neq b \neq c$	$\alpha \neq \beta \neq \gamma$
monoclinic	$a \neq b \neq c$	$\alpha = \gamma = 90° \neq \beta$
orthorhombic	$a \neq b \neq c$	$\alpha = \beta = \gamma = 90°$
tetragonal	$a = b \neq c$	$\alpha = \beta = \gamma = 90°$
rhombohedral	$a = b = c$	$\alpha = \beta = \gamma \neq 90°$
hexagonal	$a = b \neq c$	$\alpha = \beta = 90°; \gamma = 120°$
cubic	$a = b = c$	$\alpha = \beta = \gamma = 90°$

Typical values for the length of a unit cell axis for small molecules lie in a range of 5 to 30 Å.
Since the number of intensity peaks observed at a given orientation of the crystal is of finite size, as easily depictable by the Ewald construction mentioned in section 1.1.2 and reference [11], one needs to move the crystal during the measurement and observe the change of intensity on the detector with regard to the movement of the crystal. To elucidate further schematic representation of the Ewald construction is shown in Figure 2.3. The length of vector S_0 is proportional to the inverse of the wavelength λ^{-1} and dictates the size of the sphere. The reciprocal lattice of the crystal has its origin in the center of the crystal position, which is within the X-ray primary beam. The Ewald sphere is constructed in the opposite direction of the beam for convenience in this representation. The spacing and angles of the reciprocal lattice points is determined by the unit cell and the direction of the reciprocal lattice is given by the orientation of the crystal. If any reciprocal lattice point (assuming it has a finite size due to beam size, thermal motion etc.) lies in the surface of the Ewald sphere a vector S, connecting the origin of the sphere and the lattice point can be constructed that gives the vector of the diffracted beam starting at the crystal position. The angle between the two vectors S and S_0 is the angle between the incident beam and the diffracted beam, which would be required to observe this reflection. The difference vector, originating at the crystal position, between S_0 and S is the diffraction vector h. This vector needs to be an integer linear combination of the three reciprocal space vectors of the unit cell. The integer numbers for this

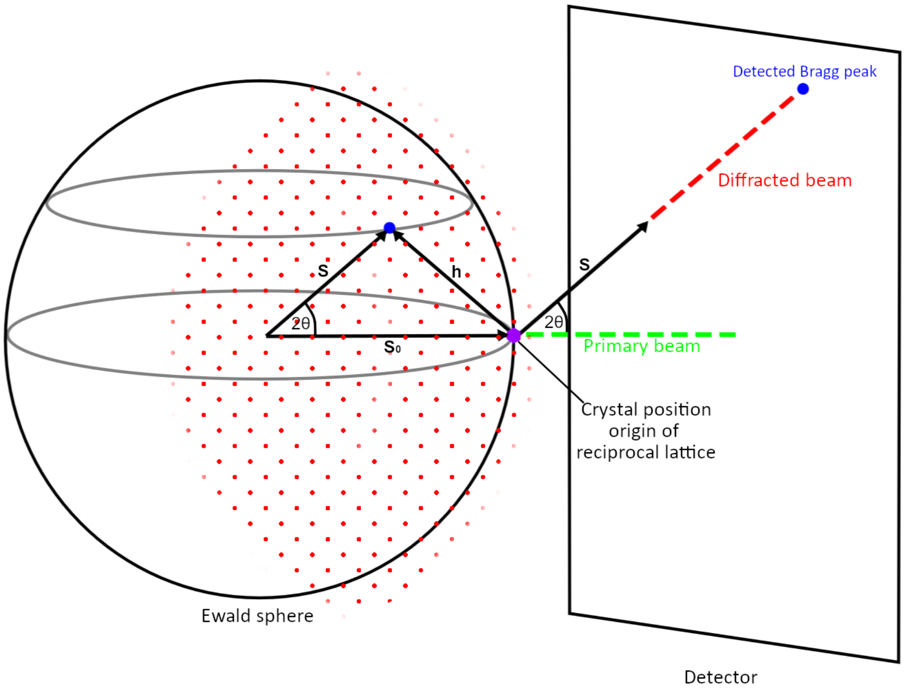


Figure 2.3: Schematic representation of the Ewald construction [11] showing the Ewald sphere, crystal and reciprocal space origin position (purple), reciprocal lattice points (red dots), the incident radiation beam (green dashed line), a selected reciprocal lattice point, that will show as a Bragg peak on the detector (blue point) and the corresponding diffracted beam (red dashed line). The picture was drawn by and reprinted with permission of Stella Geiger.

combination are called *Miller* indices h, k, l and correspond to lattice planes intersecting the unit cell at a/h,b/k and c/l. [322]

Employing these relations for a small number of measured Bragg peaks it is possible not only to determine the unit cell through the evaluation of the distance between reciprocal lattice points, but also to find the orientation matrix linking the crystal axis system with the lab axis system of the goniometer. It should be mentioned that one can also construct the Ewald sphere around the crystal position, but the presented setting, describing the same connection between angles and vectors, was chosen for clarity reasons.

Integration of Intensities

The observed intensities on a detector for those pixels, which are attributed to one reflection in the experiment taking into account the size of reflection peaks, are integrated over the area covered by the reflection, as well as over several frames of a measurement. This procedure can either be done by assuming a box for the integration in which intensity data is just summed up around the expected position of a reflection based on the Ewald construction in all dimensions (2D on the detector image and the third dimension being the consecutive frames). The more sophisticated approach follows the intensities pixel per pixel and decides what the shape of the reflection is. The integration is then carried out using this "peak-profile".

This raw intensity data is usually treated for the background of the measurement by following the intensity distribution in regions where no intensity is expected, as well as attributing for dark current, which is present in the case of CCD detectors (see also section 1.1.4). Hybrid pixel detectors suffer less form these effects by their very concept and construction.

Based on the level of noise, background, and shape consistency of the measured signal the uncertainty of an integrated intensity is evaluated, and the standard uncertainty is saved together with the intensity data. Additionally, the differences in intensity due to different exposure times or angular motion speed is accounted for by relative scaling of individual data to bring them on the same scale. In the end, the amount of information is reduced drastically from several hundreds of frames to a list of reflections, which are sorted by their respective

Miller indices. [322]

Data Reduction

The amount of data can be reduced further through the knowledge of the space group of a structure and therefore the symmetric relations in the measured reflection pattern. If there are symmetric patterns in the unit cell of a crystal structure, e.g. mirror planes, screws, glides or inversion centers, they reduce the amount of needed diffraction pattern to have complete information about a crystal. This is the case since the diffraction pattern will follow these symmetric relations in terms of identical intensities of reflections which correspond to symmetry-related miller indices of this space group. The different classes of symmetry relations in the diffraction pattern are categorized by the *Laue* classes. [116] These symmetry-related *Bragg* peaks will have identical intensity.
This is true except for an effect in non-centrosymmetric space groups, which is called anomalous dispersion that leads to a difference in phase and intensity of related reflections. This is due to the difference in interaction with the electron density in the crystal based on the path the X-ray follows through the crystal. The scattering of incident X-rays is in reality not purely elastic. The inelastic scattering is due to partial excitation of electrons in the atoms that is not directly released in the same way, leading to a shift in wavelength and phase of the emitted photon. In centrosymmetric space groups this effect cancels out in terms of the phase shift, while the complete effect of anomalous dispersion can be observed in the case of non-centrosymmetric space groups. Since there is chiral information contained in the crystal electron density in these cases, the diffraction pattern represents this difference of due to the inelastic part of the diffraction. This effect is most severe for wavelengths of X-rays which are close to the absorption edges of one of the atom types present in the crystal.
Based on this knowledge about symmetry relations between reflections it is possible to merge reflections. These reflections might have appeared on different positions of the detector and have different miller indices by direct indexing. It is required to apply the corresponding symmetry matrices to their indices to match them and to find a scale factors for the measured intensity, if for example different intensity or background levels etc. are observed. This is especially needed if the reflection was measured with different X-ray intensities, different exposure times of the detector, or by a different path of the beam through the crystal to treat for absorption effects of the crystal, as there is always a certain amount of light absorbed by the crystal due to the inelastic interaction between photons and matter. These effects and other error models like a change in beam intensity, beam profile, and effects like incident beam correction are taken into account and the amount of data is ultimately reduced to the list of information about the reflections needed to describe the whole diffraction pattern.

2.1.4 Structure Solution

To interpret the measured intensities a structure factor $F_{\vec{h}}$ is required, which can be generated from a model of the structure by describing its density and calculating its Fourier transform with regard to the cell axis system according to: [323]

$$F_{\vec{h}} = \int \rho(r)\,\mathrm{d}V \cdot e^{2\pi i \vec{h}\cdot\vec{r}} \cdot e^{T} \tag{2.2}$$

$$\approx \sum_{j}^{N} f_j \cdot e^{2\pi i \vec{h}\cdot\vec{r}} \cdot e^{T_j}. \tag{2.3}$$

Here N denotes the number of atoms in the asymmetric unit of the unit cell and f_j the atomic scattering factor of atom j, which is a model to partition the integral of electron density into atomic contributions. This model assumption is however not necessarily correct and the level of assumption highly depends on the selected atomic model. The advantage of this approach is the reduction of complexity of having a continuous, not necessarily analytically known, electron density to work with, into atomic contributions. This also allows for the atomic displacement to be separated into atomic contributions. The atomic scattering or form factors can be generated in different ways (see sections 2.1.6, 2.1.7 and 2.1.8). T_j is, in first approximation, the description of the thermal movement of the atom, which is usually done in the form of ADPs. ADPs describe the motion of atoms and their displacement throughout the crystal lattice from their respective equilibrium distance. They can either be isotropic being a single value U_{iso} or a higher tensor matrix, called U_{ij} in the case of harmonic anisotropic movement, with

$$T = -2\pi^2 u_j^2 d_{\vec{h}}^{-2}, \tag{2.4}$$

where u_j^2 corresponds to the displacement parameter of atom j and $d_{\vec{h}}$ denotes the d-spacing of a selected miller index $\vec{h}$, since the effect of atomic discplacement is depending on the scattering vector. [322] In the case of anisotropic atom displacement the expression becomes

$$T = -2\pi^2 (U_{11}h^2 a^{*^2} + U_{22}k^2 b^{*^2} + U_{33}l^2 c^{*^2} + 2U_{23}klb^*c^* + 2U_{13}hla^*c^* + 2U_{12}hka^*b^*). \tag{2.5}$$

If the factor of 2π is included and u converted from the d-spacing into $\frac{sin^2\theta}{\lambda^2}$ then one calls this the *Debye-Waller* factor B. [324, 325]
Unfortunately, having the reduced data from the experiment alone, is insufficient to *a priori* calculate or see the atomistic structure. There is a lack of a crucial piece of information for any structural model of crystallography to work. The measured intensity is only proportional to the square of the structure factor:

$$I_{\vec{h}} \approx \left|F_{\vec{h}}\right|^2. \tag{2.6}$$

This leads to the *phase problem of crystallography*, which consists of the absence of information about the phase (ϕ) of the structure factor $F_{\vec{h}}$ in the complex plane. It corresponds to the value of the scalar product of the Miller vector $\vec{h}$ of the reflection and the position in space for the corresponding electron density $\vec{r}$. Therefore the assignment of atom positions in the unit cell could in principle be ambiguous with regard to a corresponding *Miller* index. This is the case since the relative distances between atoms is preserved in the intensity information, but their absolute position in the unit cell with respect to the origin is depending on the phase of the structure factor. Since the measured intensities are proportional to the square of $F(\vec{h})$ the problem of determination of the phase, that is defining the orientation of scattering vectors in the complex plane which ultimately refers to the index in reciprocal space is necessary in order to know the true atomic structure of molecules in the crystal.
There are different approaches to solve this problem. One of them is based on the fact that the relative distance and the relative height of electron density close to the position of atoms is independent of the phase of the structure factors. By employing the so-called *Patterson* function, which depends on the square of the structure factor a corresponding *Patterson* map can be generated, which shows peaks at positions relative to the origin which correspond to the distance between two atoms. [21] While the peak at the origin is proportional to the square of the number of electrons in the unit cell the other peaks are proportional to the product of the atomic number of the elements, which makes it easier to distinguish distances between heavy elements from lighter elements. Also structural motifs can be recognized from the *Patterson* map, taking into account that for example a six-membered ring will show as a superposition of six rings, where each ring is shifted in the map around the origin placing each of the centers in the origin. Using this knowledge algorithms can make an assumption of the three-dimensional structure of the crystal and calculate phases based on this model for further refinement.
A different approach is the *direct methods* solution. For space groups containing an inversion center the phase problem is less complicated, as only a shift of π in the phase of $F_{\vec{h}}$ is allowed, as required by the symmetry. In non-centrosymmetric space groups, the problem is more severe, but in the case of many reflections determined in the experiment it is possible to use for example brute force approaches, that is using random phases and evaluating the resulting density obtained by Fourier synthesis, to solve the phase, since only one phase would result in reasonable electron density maps, taking into account knowledge about electron density distributions around atoms. This procedure is repeated until a satisfactory electron distribution is found. Hence the name *multi solution techniques*. [21]
In case of centrosymmetric space groups the solution of a phase can be achieved by assigning random values of 0 and π to three selected intense reflections. A reflection with a phase of 0 is assigned a sign based on their sign of the real part in the complex plane when assigning the angle to unity: $S_{\vec{h}} = +$, a reflection with phase π

is assigned a negative sign $S_{\vec{h}} = -$. If the indices are selected in a way that reflections have indices $\vec{h}$, $\vec{h'}$ and $\vec{h} - \vec{h'}$ *Sayre's* [326] equation becomes true:

$$S_{\vec{h}} = S_{\vec{h'}} \cdot S_{\vec{h}-\vec{h'}}. \tag{2.7}$$

In non-centrosymmetric space groups the *Sayre* equation becomes more general:

$$F_{\vec{h}} = \sum_{\vec{h'}} F_{\vec{h'}} F_{\vec{h}-\vec{h'}} \tag{2.8}$$

In the original publication this equation was constructed for monoatomic crystals but stated, that it "[...]will be applicable to the entire class of ordinary organic crystals [...] not covered by the heavy-atom technique.". [326] If the equations 2.7 or 2.8, depending on the space group, are fulfilled, it is probable, that a set of determined phases is correct. This approximation by *Sayre* is based on the assumption of monoatomic crystals, which can be transferred to light atoms. For a more general procedure for non centrosymmetric space groups the method by *Karle* and *Hauptmann* is used. [327] Here the relation of phases can be expressed as:

$$\phi_{\vec{h}} = \phi_{\vec{h'}} + \phi_{\vec{h}-\vec{h'}}. \tag{2.9}$$

Employing the tangent formula they deduct: [21, 328]

$$tan\phi_{\vec{h}} = \frac{\sum_{\vec{h}} \kappa \cdot sin(\phi_{\vec{h'}} + \phi_{\vec{h}-\vec{h'}})}{\sum_{\vec{h}} \kappa \cdot cos(\phi_{\vec{h'}} + \phi_{\vec{h}-\vec{h'}})}, \tag{2.10}$$

where κ introduces the probability of having correct phases based on E statistics, by $\frac{1}{\sqrt{N}}|E_{\vec{h}}E_{\vec{h'}}E_{\vec{h}-\vec{h'}}|$. The E-values are obtained by the calculation of normalized structure factors using

$$E^2 = k\frac{F^2_{\vec{h}}}{F^2_{\vec{h},exp}}, \tag{2.11}$$

where k is a fixed scale factor scaling the measured intensities to the number of electrons expected in the unit cell. $F^2_{\vec{h},exp}$ is the expected structure factors, taking into account the knowledge about the resolution dependent decay of structure factors at higher angles using the scattering factors of atoms f_j, expressed by

$$F^2_{\vec{h},exp} = \epsilon \sum f_j^2, \tag{2.12}$$

where ϵ denotes a small integer number necessary for some classes of reflections due to the crystal lattice. If these values of E are $>> 1$ or $<< 1$ the distinction between high intensity reflections and low intensity reflections becomes possible and allows selection of reflections that have very good signal to noise ratio for the calculation of phases.

One of the most successful approaches for structure solution in modern software, like for example *SHELXT*, [329] is called dual-space methods. They make use of *Patterson* methods and combine them with the knowledge about the direct methods and the corresponding relations of phases to obtain the correct solution by applying masks to the structure factors to iteratively remove bits of the known solution, allowing the search other missing structural motives on the remaining information. Subsequently the model is improved by assumptions for scale factors to match typical peaks in the density that could correspond to carbon atoms and heavy elements to correspond to the highest density peak of the phase solution. The solution happens in the P1 space group, which also allow the model to refine if the wrong space group was assumed during the data reduction. This method was used for all structures reported in this thesis.

Another approach is the use of the Charge-Flipping-Algorithm. [330, 331] Here a random set of phases is assumed and the density in the unit cell calculated by Fourrier synthesis. After this initial cycle all regions in the unit cell that show negative electron density are "flipped", that is the sign of the density reverted (hence the

name), and new structure factors and their phases calculated from this new density. This is repeated in a cyclical manner until the changes converge.
After the solution of phases, the positions of points in the unit cell with high electron densities are defined, which can be assigned to atomic positions, based on the value of electron density measured. With the knowledge of phases, the structure needs to be refined and completed. This includes the correct assignment of positions of atoms to peaks in the electron density map obtained from Fourier transforming the measured structure factors, the adequate description of their ADPs and in the case of non-spherical structure factors a correct description of the non-sphericity and contribution to the structure factor of the atoms.

2.1.5 Least-squares Refinement

In a crystallographic Least-squares (LS) refinement, as for example implemented in *olex2.refine*, [50] the minimization of a weighted square of differences between measured and calculated structure factors is performed. This is achieved by minimizing the LS function

$$LS(\vec{x},K) = \underbrace{\sum_{\vec{h}} w_{\vec{h}} \left(F^o_{\vec{h}} - KF^c_{\vec{h}}(\vec{x})\right)^2}_{\text{data}} + \underbrace{\sum_{restraint\ i} w_i \left(T^t_i - T^c_i\right)^2}_{\text{restraints}}, \tag{2.13}$$

where F^o and F^c are the observed and calculated structure factors, T^t and T^c are target and current model restraint parameters, K an unknown scale factor bringing measured and calculated data onto the same scale and w a weight for each reflection. This weight is in many cases based on measured uncertainties $\sigma_{\vec{h}}(F^o)$ in a statistical relation $w = 1/\sigma^2(F^o)$, in more complex weighting schemes also on F^o, F^c and $\vec{h}$ themselves. If there are no constraints or restraints applied the second term vanishes. $\vec{x}$ describes a vector of atomic positions, ADPs and occupancies used in the refinement.
In a first step a minimization of LS is performed changing only K, keeping $\vec{x}$ fixed. Subsequently a set of substitutions is performed, replacing the restrained or constrained parameters of $\vec{x}$ by their known analytical dependency on the parameters of $\vec{x}$. Then the derivatives of these substituted equations for $\widetilde{F}^c$ and $\widetilde{T}^c$ with regard to all remaining parameters in $\vec{x}$ are calculated. This means the influence of a change in parameters of $\vec{x}$ onto the values of $\widetilde{T}^c$ is calculated in addition to their direct influence through the data part of equation 2.13. Then the value of LS evaluated and subsequently minimized using the calculated gradients of the parameter vector $\vec{x}$.
In order to have reasonable scales between measured data and restraints applied, a definition of w_i is desirable, that allows similar strength of influence between residual density between model and observation and restraints. *Rollett* and *Ford* proposed the normalization factor χ^2, [332] whose square root is also known as *goodness of fit* (S), [21] to be used as a scale for restraints, which is defined as:

$$\chi^2 = w_i = \frac{1}{N_r - N_p} \sum_{\vec{h}} w_{\vec{h}} \left(F^o_{\vec{h}} - KF^c_{\vec{h}(\vec{x})}\right)^2, \tag{2.14}$$

where N_r and N_p correspond to the number of reflections and parameters, in this case the length of $\vec{h}$ and the restraint reduced $\vec{x}$.
χ^2 gives insight into the general agreement between experiment and wavefunction used to model it. If the weight function is solely relying on the uncertainty of a reflection $\sigma_{\vec{h}}$ an agreement within one σ of the measured data would result in $\chi^2 = 1$, smaller values might hint towards overestimated uncertainties, higher values indicate significant differences quite sensitively. A value of $\chi^2 = 0$ would correspond to perfect agreement, which in case of experimental data is unlikely to happen, as this would mean there would be no noise in the data and no other sources of model inaccuracy, but for structure factors calculated from a theoretical model without errors this value might become very close to 0.
The most common metrics to assess the quality of an refinement model is a combination of this goodness-of-fit,

the minima and maxima of the residual density, which is the Fourrier synthesis of the difference in structure factors and the residual (R)-statics, that are based on the average disagreement of reflections: [21]

$$R = \frac{\sum ||F_o| - |F_c||}{\sum |F_o|} \tag{2.15}$$

$$wR = \left[\frac{\sum w_{\vec{h}} \left(|F_o| - |F_c|\right)^2}{\sum w_{\vec{h}} \left(F_o\right)^2}\right]^{1/2} \tag{2.16}$$

$$wR_2 = \left[\frac{\sum w_{\vec{h}} \left(F_o^2 - F_c^2\right)^2}{\sum w_{\vec{h}} \left(F_o^2\right)^2}\right]^{1/2} . \tag{2.17}$$

The better the model the lower the R-values and the closer S, being the square root of χ^2 gets to 1. Since experimental data always contains random noise a reasonable value for R in a good systematic-error-free model of the structure is below 0.05.

2.1.6 Independent Atom Model (IAM)

In the IAM the electron density used to formulate scattering factors f is obtained by single atom wavefunction calculations, as done in the database of *Cromer* and *Mann* [41] or *Su* and *Coppens*. [333] These scattering factors represent spherical atoms, as they are the Fourier transform of isolated single atoms. Usually slater type functions are used for the calculation of these densities (see equations 2.33 - 2.36). Since these form factors do not vary with respect to the structure they are used for, they can be stored in a databank and used in equation 2.3 to build structure factors. This can be done regardless of the structure at hand or constrains onto the behaviour of an atom due to crystallographic symmetry elements present at an atomic position, since the spherical density obeys every possible symmetry. With these constructed structure factors the model is ready to be refined against the measured intensities through a least squares refinement of positions, ADPs and in cases of disordered or partially occupied structures their respective occupations (see section 2.1.5).

2.1.7 Multipole Model

Due to the shortcomings of the IAM (see section 1.1.7) the idea of an extension of the IAM emerged to account for the known and observed non-spherical behaviour of atoms, also observable in X-ray diffraction experiments. [49, 334] Development of generalized functions to describe these non-spherical densities and their application for interpretation of the measured intensities by *Stewart et al.* started already 50 years ago. [335–339] Ultimately *Hansen* and *Coppens* used the non-spherical approach to enhance the formalism of the IAM introducing the their own proposal of a Multipole Model. [66] In their formulation the extension is based on the distinction between a core and valence contribution to the electron density of an atom, that is expressed in a series of radial function in combination with spherical harmonics (compare equation 2.35). This way f_j from equation 2.3 is not spherical and can take into account bonding and deformation of the atom due to environmental effects.

$$\rho_j = P_c\, \rho_{core,j}(r) + P_v\, \kappa'^3\, \rho_{valence,j}(\kappa' r) + \sum_{l=0}^{l_{max}} \kappa''^3 R_l(\kappa'' r) \sum_{m=-l}^{l} P_{lm} Y_l^{(m)}(\theta, \phi) \tag{2.18}$$

Further details of this model are not given, since the model was not used throughout this thesis and is only shortly mentioned for purposes of reference of older models.
Examples of software for this kind of refinement include *XD* [87] or *MoPro*. [88]

2.1.8 Hirshfeld Atom Refinement (HAR)

Abandoning the idea of using predefined tabulated scattering factors, as in the IAM or Multipole Model, HAR uses the idea of tailor-made scattering factors from wavefunctions calculated for each structure under investigation. The density of a molecule could in principle be of any origin. In the procedure defined by *Jayatilaka, Grabowsky et al.* [125, 140] a wavefunction is calculated for a molecular fragment. It is partitioned into atomic contributions according to the *Hirshfeld* stockholder partitioning scheme [131]

$$\omega_j(\vec{r}) = \frac{\rho_j^0(\vec{r} - \vec{r_j})}{\sum_k \rho_k^0(\vec{r} - \vec{r_k})}, \tag{2.19}$$

where each atom j has a corresponding weight function ω_j at each point in space $\vec{r}$. This weight is assigned by calculating the promolecular density for all atoms, which is a spherical electron density similar to the IAM (see section 2.1.6) but calculated using the basis set and level of theory used for the wavefunction calculation for consistency. The weight function is the ratio of electron density contributed by the promolecule atom j to the total density generated by all atoms k, with $k = j$ included. This procedure yields the atomic electron density of atom j based on the total electron density calculated from any given wavefunction with MO coefficients C:

$$\rho_j(\vec{r}, C) = \omega_j(\vec{r}) \cdot \rho_{molecule}(\vec{r}, C) \tag{2.20}$$

Convoluted with a probability distribution function $P_j(\vec{r})$, being a Fourier transform of the thermal term e^T in equation 2.3, the description of a thermally smeared *Hirshfeld*-atom density $\langle \rho_j \rangle$ is expressed as:

$$\langle \rho_j(\vec{r}, C) \rangle = \rho_j(\vec{r}, C) * P_j(\vec{r}) \tag{2.21}$$

where $*$ represents the convolution. These thermally smeared electron densities can be used in the same manner as in the multipole model to give structure factors similar to equation 2.3 by constructing structure factors from the *Hirshfeld*-atoms:

$$F_{\vec{h}} = \sum_j \int \langle \rho_j(\vec{r}, C) \rangle \cdot e^{2\pi i \vec{h} \cdot \vec{r}} \mathrm{d}V \tag{2.22}$$

Since the calculation of a wavefunction and the evaluation of its electron density in the whole asymmetric unit or unit cell is computationally demanding the properties of *Becke*-grids [340] are employed to evaluate the electron density using radial and angular quadrature procedures of *Gauss-Chebyshev* second type and *Lebedev* angular quadrature of the unit sphere. [341] This reduces the number of points to be evaluated significantly in comparison to equidistant evaluation of volumetric data (see section 1.3.1).

These scattering factors are used in the least squares refinement procedure (see section 2.1.5). The sequence of wavefunction calculation, partitioning, calculation of structure factors and least squares fit is called a rigid body fit. During this procedure the atomic shape is considered to be rigid and independent of atom displacement, that means the form factor f_j is considered to be constant throughout the least squares procedure. If this procedure is repeated iteratively until a convergence criteria in atomic positions and ADPs is achieved it is called *Hirshfeld* Atom Refinement (HAR).

In this type of refinement the purpose is to use the better scattering factors obtained by Hirshfeld partitioning in order to obtain better structural models of the compounds under investigation. The MO coefficients C are not directly modeled, but a result of a normal wavefunction calculation. This clearly makes HAR a pure structural refinement technique, which fulfills the first aspect of Quantum Crystallography, as defined in section 1.2. For a comparison of these form factors in comparison to the IAM-spherical form factors see Figures 1.6 and 1.7.

2.1.9 X-ray Constrained Wavefunction Fitting (XCW)

If one wants to get a model of the electron density without changing the structural model, it is possible to use the formalism of non-spherical atoms, for example the *Hirshfeld*-atoms, to achieve this in an independent procedure. Here the atomic partitioning is only required to introduce the effect of atomic displacement parameters. Otherwise a direct Fourier transform of the symmetry generated unit cell electron density based on the single-molecule wavefunction electron density would be sufficient. By calculation of the agreement of a vector of calculated structure factors $\vec{F^c}(C)$ (see equations 2.22 & 2.3) in terms of χ^2 statistics (see equation 2.14 with

$w_{\vec{h}} = 1/\sigma^2$) between the calculated and the experimentally derived structure factors (compare equation 2.6) a measure of similarity between model and the observed data is found. This is similar to the IAM. Instead of varying the positions and thermal displacement parameters of atoms it can now be rationalized that a change in the MO-coefficients might also affect the agreement between model and observed data through the Hirshfeld-atom-formalism. The agreement becomes a function of the coefficients C, as well:

$$\chi(\vec{F}^o, C, \vec{x})^2 = \frac{1}{N_r - N_p} \sum_{\vec{h}} \frac{\left(\left|F^c_{\vec{h}}(C, \vec{x})\right| - \left|F^o_{\vec{h}}\right|\right)^2}{\sigma^2_{\vec{h}}}. \tag{2.23}$$

N_r and N_p are the number of measured unique reflections and used parameters during the refinement in terms of positions, ADPs and other variables summarized in $\vec{x}$. The term $F^c_{\vec{h}}(C, \vec{x})$ is the structure factor calculated for MO coefficients C based on this model structure (see equation 2.45 and the following paragraph) and $F^o_{\vec{h}}$, being the square root of the experimentally measured intensity and the corresponding uncertainty $\sigma_{\vec{h}}$.
Using the sensible definition of the goodness-of-fit *Jayatilaka* proposed a manually adjusted implementation of perturbation theory to describe a procedure to use a *Lagrange* (undetermined) multiplier as a tunable variable to perturb the the energy $E(C)$ with the χ^2 agreement reduced by a desired target level of agreement Δ:

$$L_{XCW}(\vec{F}^o, C, \lambda) = E(C) + \lambda \left(\chi(\vec{F}^o, C, \vec{x})^2 - \Delta\right). \tag{2.24}$$

If instead of the energy $E(C)$ the Lagrangian function L_{XCW} is used as a minimization target for a Hartree-Fock (HF)- or DFT-Self-Consistent field (SCF) procedure, the resulting wavefunction corresponds to the variational minimum according the experimentally perturbed *Fock* matrices. This means it contains experimental information to a degree directly depending on the *Lagrange* multiplier λ and is not a variational minimum to the pure SCF procedure anymore. By slowly increasing the value of λ from zero, which corresponds to the wavefunction normal *ab initio* wavefunction, to a higher value subsequently more experimental perturbation is included in the resulting wavefunction. By increasing λ, the value of χ^2 usually converges to a minimum, as shown on an example set of theoretical structure factors constructed using a correlated wavefunction which were fitted by a HF-XCW *ansatz*, in Figure 2.4, where a convergence close to $\chi^2 = 0$ is to be expected, as there is no noise present in the structure factors. The remaining difference is the incapability of the HF-*ansatz* to fully describe a correlated wavefunction.

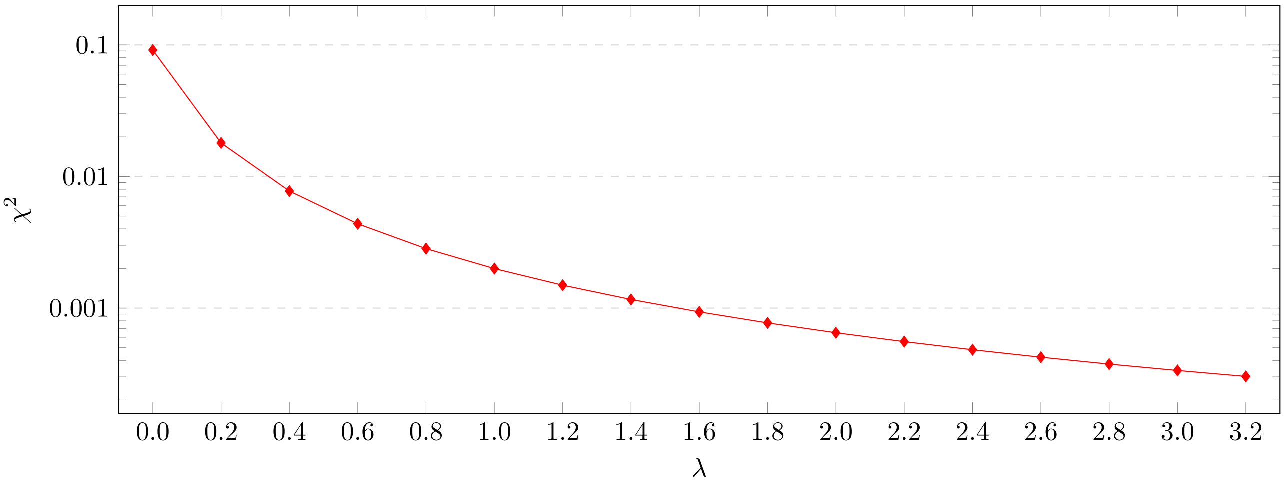


Figure 2.4: Plot of the *goodness of fit* χ^2 against the Lagrange multiplier λ used for a test set of noise free structure factors for alanine in the gas phase. The structure factors were created from a DFT calculation, and fitted using HF-XCW for the wavefunction using *Tonto*.

XCW is a technique to improve the electron density description of the structure using a modification of the wavefunction while keeping the structure itself fixed. This enhances the wavefunction based on the experimental data and hence fulfills the second aspect of Quantum Crystallography, as defined in section 1.2.

2.1.10 X-ray Wavefunction Refinement (XWR)

The Hirshfeld atom refinement (HAR) is used to refine structural parameters. The XCW method, on the other hand, improves the description of the electron density while retaining the same structural parameters. Thus, a combination of both techniques should yield the best description of the experimental data. If these steps are performed once each, this procedure is called XWR. [92] If one would want to proceed, by using the improved description of the wavefunction obtained from XCW, the next step can only be a rigid atom fit using the predefined wavefunction, as HAR would include a recalculation of an unperturbed wavefunction. If XCW and rigid atom fit are performed subsequently until convergence in energy and structure are achieved the procedure is called total X-ray Wavefunction Refinement. It depends on the point of view, whether total-XWR is to be considered a more sophisticated version of HAR or it should be considered a separate technique to reserve the term HAR solely for the unperturbed wavefunction technique and *ab initio* structural refinement without experimental data being introduced into the scattering factors. This would raise the question whether the inclusion of the MO coefficients as parameters of the model would be necessary and therefore refinement statistics would have to be redefined in this framework.

2.2 Wavefunction Calculations

A central equation in quantum mechanics is the time-independent *Schrödinger* equation:

$$\hat{H}\Psi = E\Psi. \tag{2.25}$$

This equation gives the energy E of a wavefunction Ψ when the Hamiltonian $\hat{H}$ is applied onto it. Therefore the Hamiltonian needs to be defined. It contains kinetic ($\hat{T}$) and potential ($\hat{V}$) contributions, which can be stated as:

$$\hat{H} = \hat{T} + \hat{V}. \tag{2.26}$$

The terms for the kinetic energy and the potential energy here contain all contributions, the n electrons' and N nuclei's motions and their interactions. They can be separated and expressed in the following way:

$$\begin{aligned}\hat{H} = &-\frac{\hbar^2}{2m_e}\sum_{i=1}^{n}\nabla_i^2 - \frac{\hbar^2}{2}\sum_{K=1}^{N}\frac{1}{M_K}\nabla_K^2 \\ &+\sum_{K=1}^{N}\sum_{L>K}^{N}\frac{Z_K Z_L e^2}{4\pi\epsilon_0 R_{KL}} + \sum_{i=1}^{n}\sum_{j>i}^{n}\frac{e^2}{4\pi\epsilon_0 r_{ij}} \\ &-\sum_{K=1}^{N}\sum_{i=1}^{n}\frac{Z_K e^2}{4\pi\epsilon_0 R_{Ki}}.\end{aligned} \tag{2.27}$$

The first two terms are the electron and nuclear kinetic energies. The third and fourth terms are the nuclear-nuclear and electron-electron repulsion and the fifth term is the nuclear-electron attraction due to Coulomb interaction. With this Hamiltonian it is possible to retrieve energies for a particular wavefunction. But the 4th term is problematic, because the electron-electron interaction is not analytically solvable. There needs to be a process to approximate it in the best way possible.

Regardless of any approximations made, the absolute square of the value of the wavefunction gives rise to the electron density ρ of this wavefunction at any given point in space $\vec{r}$:

$$\begin{aligned}\rho(\vec{r}) &= n\sum_{spins}\int |\Psi(\vec{r}_1,\vec{r}_2,\ldots,\vec{r}_n)|^2\,\mathrm{d}\vec{r}_2\ldots\vec{r}_n \\ &\equiv n\int |\Psi(\vec{r}_1,\vec{r}_2,\ldots,\vec{r}_n)|^2\,\mathrm{d}\tau'\end{aligned} \tag{2.28}$$

2.2.1 Approximations

The following three approximations are used in quantum chemistry to make the problem solvable:

- *Born-Oppenheimer*-approximation
- Orbital approximation
- Linear combination of Atomic Orbitals to Molecular Orbitals (LCAO-MO) approximation

Only when these approximations are considered, it is possible in the framework of quantum chemistry to solve the *Schrödinger* equation for a system containing more than one electron. The mentioned approximations will be summarized briefly.

Born-Oppenheimer approximation

The *Born-Oppenheimer* approximation neglects the movement and vibration of nuclei during the calculation of the wavefunction. Additional terms for the nuclear description are formulated to make the electron distribution independent of nuclear movements. This is possible because the movement of electrons is considered to be almost instantaneous compared to the nuclear movement, so the electrons could instantaneously respond to changes of nuclear positions. This makes it possible to separate their movement by separating the terms in the Hamiltonian into $\hat{H}_e$ and T_n:

$$\begin{aligned}\hat{H}_e =& -\frac{\hbar^2}{2m_e}\sum_{i=1}^{n}\nabla_i^2 + \sum_{K=1}^{N}\sum_{L>K}^{N}\frac{Z_K Z_L e^2}{4\pi\epsilon_0 R_{KL}} \\ &+ \sum_{i=1}^{n}\sum_{j>i}^{n}\frac{e^2}{4\pi\epsilon_0 r_{ij}} - \sum_{K=1}^{N}\sum_{i=1}^{n}\frac{Z_K e^2}{4\pi\epsilon_0 R_{Ki}}\end{aligned} \tag{2.29}$$

$$T_n = -\frac{\hbar^2}{2}\sum_{K=1}^{N}\frac{1}{M_K}\nabla_K^2 \tag{2.30}$$

This does not neglect electrostatic interaction between electrons and nuclei, but allows the deconvolution of atomic motion from electron motion. The shift of nuclear positions then does not interfere with the shape or distribution of the electrons and can be evaluated independently, being reintroduced later.

Orbital approximation

This approximation postulates the movement of each electron to be separable of the movement of others. Therefore their function can be separated into a series of independent multiplicands, which then can be described independently. By employing this product assumption, the wavefunction of a molecule becomes a product of contributions of every single electron. This postulate is based on the observation that the character and behavior of an electron is mostly depending on the energetic level it is located in. Valence electrons participate in bonds, core electrons do not. Therefore individual energetic levels are separated. This makes it easier to take the sixth postulate of quantum theory into account, the so-called *Pauli*-principle, which requires the wavefunction to change sign when two electrons are swapped. This is due to the fermionic character of the electron. The orbital approximation allows all single electron functions to be put into the *Slater*-determinant. Such a determinant intrinsically ensures the change of sign of the wavefunction when two rows or two columns are interchanged, and that it vanishes when two rows or columns are equal. Doing so results in the wavefunction of an n electron system as the following *Slater*-determinant:

$$\Psi_e = \frac{1}{\sqrt{n!}}\begin{vmatrix}\phi_1(1) & \phi_2(1) & ... & \phi_n(1)\\ \phi_1(2) & \phi_2(2) & ... & \phi_n(2)\\ : & : & : & :\\ \phi_1(n) & \phi_2(n) & ... & \phi_n(n)\end{vmatrix} \tag{2.31}$$

Here ϕ_i is the i^{th} MO of the molecule.

LCAO-MO approximation

The *Born-Oppenheimer* and the orbital approximations are still insufficient to solve the *Schrödinger* equation. Another approximation is necessary: MOs are assumed to be linear combinations of Atomic Orbitals (AOs). Each MO now is only a sum of weighted AOs.

$$\phi_i = \sum_{\mu} c_{\mu i} \chi_{\mu} \tag{2.32}$$

Here ϕ_i is a MO, χ_μ an AO and $c_{\mu i}$ is the MO-coefficient. The mathematical description of AOs is pre-calculated and stored in tables, the so called basis sets (compare 2.2.2).
This yields a matrix of coefficients $C_{\mu i}$ for the contribution of each AO to the MO. The product of this matrix and a vector containing all AOs results in the MO.
The two most common types of functions used for AOs are *Slater*- and Gaussian-type functions. The *Slater*-type function is defined as:

$$\chi_{Slater}(\vec{r}) = R_{Slater,l}(r) Y_l^{(m)}(\theta, \phi) \tag{2.33}$$

where $R_{Slater,n}(r)$ refers to the radial *Slater*-type function for principal quantum number n:

$$R_{Slater,n}(r) = r^{n-1} e^{-\zeta r} \, (2\zeta)^n \sqrt{\frac{2\zeta}{(2n)!}}. \tag{2.34}$$

$Y_l^{(m)}(\theta, \phi)$ are the spherical harmonics with angular momenti l and m:

$$Y_l^{(m)}(\theta, \phi) = \sqrt{\frac{2l+1}{4\pi} \frac{(l-m)!}{(l+m)!}} P_l^{(m)}(cos\theta) e^{im\phi} \tag{2.35}$$

with $P_l^{(m)}$ referring to the associated *Legandre* polynomials:

$$P_l^{(m)}(x) = \frac{(-1)^m}{2^l l!} \left(1 - x^2\right)^{m/2} \frac{\mathrm{d}^{l+m}}{\mathrm{d}x^{l+m}} \left(x^2 - 1\right)^l \tag{2.36}$$

Slater-type functions are closer to the experimentally measurable electron density of an atom. They are capable of representing the cusp at the nucleus and more diffuse behavior at higher distance, but because of their cusp at the atomic nucleus it is hard and computationally costly to integrate and calculate MOs with it. Therefore they are not the ideal choice for quantum chemical calculations.
The Gaussian-type function is mostly used, which is easy to integrate and easily combinable with others. The disadvantage is some discrepancy from the electron density measured. The decay is too steep further away from the nuclear positions and there is no cusp at the nuclear position. This can be corrected by taking more than one Gaussian-type function. A function containing more than one Gaussian-type function can also be expressed as a Gaussian-type function with different factor and exponent. A Gaussian-type function can be expressed in a similar way to *Slater*-type functions:

$$\chi_{Gauss}(r) = R_{Gauss,l}(r) Y_l^{(m)}(\theta, \phi) \tag{2.37}$$

where $Y_l^{(m)}(\theta, \phi)$ refers to the spherical harmonics (see equation 2.35) while the radial part $R_{Gauss,n}(r)$ is defined as:

$$R_{Gauss,l}(r) = B(l, \zeta) r^l e^{-\zeta \, r^2} \tag{2.38}$$

$B(n, \zeta)$ is the normalisation constant derived from:

$$\int_0^\infty r^2 \left| R_n(r) \right|^2 \mathrm{d}r = 1 \tag{2.39}$$

Contracted AOs then result using:

$$R_l(r) = r^l \sum_{p=1,P} c_p B\left(l, \zeta_p\right) e^{-\zeta_p r^2} \tag{2.40}$$

Where c_p is the contraction coefficient for a primitive Gaussian-type function with exponent ζ_p. This yields the contracted AO $R_l(r)$. Usually these are used in the wavefunction evaluation procedure and for expressing the AOs in the LCAO-MO approximation.

In summary the electronic wavefunction is a normalized determinant of MOs, which are a weighted sum of AOs, themselves being a weighted sum of Gaussian-type function.

2.2.2 Basis Sets

The AOs created this way are called a basis set. They are documented for different atoms and in different complexity, corresponding to the level of sophistication one wants to employ. Furthermore they were documented in a standardized way, so many programs can use them. Consisting of the number, weighting factor, radial decay and form for the Gaussian-type functions they do not need to be calculated during the quantum-chemical calculation. Only the MO coefficient matrix $C_{\mu i}$ needs to be calculated. The so called *Pople* basis sets [342] have the naming scheme:

$$C - V_1 V_2 V_3 (++) G(xp, yd, zf) \tag{2.41}$$

Here C denotes the number of Gaussian-type functions used to model the core of the atom. V_1, V_2 and V_3 are the number of Gaussian-type functions used for the valence shell. Typical values for V_1 range between two and three. V_2 can range between one and two and V_3 usually occurs with a value of one only for high basis sets. The presence of these numbers is referred to as the number of ζ type function. Each of these shells only consists of a series of contracted basis functions (see equation 2.40). $+$ symbols denote the use and number of diffuse functions, required for long range interactions due to their low decay at high distances achieved by comparably small ζ values. x, y and z denote the use of polarisation functions to polarize the valence shell by mixing together partial character of p-,d-,f or even g-orbitals for non-hydrogen atoms. This changes the shape of valence orbitals, so they can give better energies and interactions in the molecule and environment. For hydrogen usually there is only p or d-polarization.

The *Pople*-basis-sets [342] are very common but when it comes to more precise results usually *Dunning* [343] or *Ahlrichs* [344, 345] basis sets are used.

Dunning basis sets have a naming scheme which consists of a number corresponding to the complexity of the basis set in general and pre- and postfixes, which indicate additional features. The general form is VNZ, where N=D,T,Q,5,6 (double, triple, quadruple...). A 'cc-p' prefix indicates 'correlation-consistent polarized' basis set, 'aug' denotes augmentation through additional diffuse functions and V indicates the valence-only description of the basis sets. These basis sets are valid for the first and second period elements. For third period elements additional functions are needed, resulting in basis sets called cc-pV(N+d)Z.

The def2 basis sets, used in the scope of this work, are polarized basis sets which give good results for elements all over the periodic table, so they can be widely used and give similar results for all calculations. They were proposed by *Ahlrichs et al.*. [344]

Pollak and *Weigend* published a very valuable new set of basis sets, [154] the so called x2c-family, that is optimized for the use with relativistic Hamiltonians, especially for one and two component approximations. Their basis sets are based on the def2-family and are defined as all electron basis sets for element up to Z=86 (Rn). The def2-basis sets are defined up to Kr with all electron basis functions, but rely on the use of effective core potentials to mimic the inner shells of heavier element to reduce computational costs of these elements, as all-electron basis sets for these elements easily become computationally extraordinarily demanding. However, these effective core potentials are approximations, that might not always be desired in the course of calculations, e.g. if the presence of the core electrons is required for later analysis like the calculation of scattering factors and the filling of the core by substitution of an effective core potential might be too crude of an approximation or simply not available. In the analysis of the electron density for Quantum Theory of Atoms in Molecules (QTAIM) it is possible to augment the calculated electronical structure employing effective core potentials by a series of spherical Gaussian-type functions. [346–348] Unfortunately, this functionality is not implemented in HAR and also not directly accessible in some software. Also care must be taken which level of approximation is chosen, since it was shown in reference [347] that large-core potentials might lead to less significantly accurate results compared to all-electron calculations.

A similar contribution to make the whole periodic table of elements available for all-electron quantum mechanical calculations is the *Jorge*-family of basis sets. [155, 156] They are defined up to Z=103 and therefore can describe all atom types which might occur in long lived materials. Heavier atom types are not stable enough to be characterized in a way to allow precise and accurate parametrization of basis sets. The *Jorge* basis sets are available in double-ζ and triple-ζ accuracy and in both relativistic contracted and classical Hamiltonian optimized form. The functions were optimized for the Douglas-Kroll-Hess formulation of a two-component relativistic Hamiltonian. This allows almost all combinations of theoretical methods to be used across the complete periodic table of the elements.

2.2.3 Methods

The *method* of a calculation is the mathematical *ansatz* employed to obtain the MO-coefficients $c_{\mu i}$. The combination of method and basis set is then called *level of theory*.

2.2.3.1 Self-Consistent Field (SCF)

For SCF-calculations, the approximation is made that a selected electron interacts with the field experienced from all other electrons in their orbitals. This means the potential of an electron interacting with all other charged particles in the molecule can be expressed as:

$$V_{eff}(\vec{r_1}) = -\sum_{k}^{N} \frac{Z_k e^2}{4\pi\epsilon_0 \left|\vec{r_1} - \vec{r_k}\right|} + \sum_{i}^{n} \left(\int_V \frac{e^2\, \phi_i^*(\vec{r_i})\, \phi_i(\vec{r_i})}{4\pi\epsilon_0 \left|\vec{r_1} - \vec{r_i}\right|} \,\mathrm{d}V \right) \tag{2.42}$$

This means the charge of an electron i is already distributed in MO ϕ_i, when a first calculation is initialized. The necessity to have a so called *initial guess* arises. To obtain a starting point of coefficients for the AOs to form first MOs often a *Hückel*-guess might be used. Now the MO will be iteratively optimized in terms of energy by varying the coefficient matrix until a preselected convergence criterion is achieved. When one MO has been optimized the next one will be optimized with respect to the new electron distribution resulting from the new set of MOs from the step before. This cycling will be iterated until the energy and MO-matrix of the whole system converges, which means a self consistent electric field is experienced by all electrons.

2.2.3.2 Hartree-Fock (HF)

The variational theorem states that it is impossible to find an eigenfunction for the Hamiltonian which has lower energy than the real solution for the system. This allows the search for a solution with the lowest calculated energy which then can be called the best solution. To obtain such a solution various approaches can be used.

The most basic type of calculations is done through the HF method. Because of the formulation of the SCF the equations for a wavefunction can be separated into a system of equations for each electron and its corresponding orbital, since they are considered depending only on the average distribution of all other electrons according to their respective orbitals. The system of equations is built by:

$$\begin{aligned}\hat{F}\phi_1(1) =& \epsilon_1\phi_1(1)\\ \hat{F}\phi_2(1) =& \epsilon_2\phi_2(1)\\ \hat{F}\phi_3(1) =& \epsilon_3\phi_3(1)\\ &\vdots\\ \hat{F}\phi_n(1) =& \epsilon_n\phi_n(1).\end{aligned} \tag{2.43}$$

Here $\hat{F}$ is the *Fock* operator, given by:

$$\begin{aligned}\hat{F}_i =& -\frac{\hbar^2}{2m_e}\Delta_{el} - \frac{Ze^2}{4\pi\epsilon_0 r_{K1}}\\ &+ 2\sum_{i=1}^{n}\int \phi_i^2(2)\frac{e^2}{4\pi\epsilon_0 r_{12}}\mathrm{d}V\\ &-\sum_{i=1}^{n}\int \phi_i(2)\phi_j(2)\frac{e^2}{4\pi\epsilon_0 r_{12}}\mathrm{d}V\frac{\phi_i(1)}{\phi_j(1)}.\end{aligned} \tag{2.44}$$

Using the LCAO-MO approach to formulate the MOs out of AOs gives the *Rothaan* equations for each electron i, where μ refers to all electrons:

$$\sum_{\mu} c_{\mu i}\int \chi_i^* \hat{F}_i \chi_\mu \,\mathrm{d}V = \epsilon_i \sum_{\mu} c_{\mu i}\int \chi_i^* \chi_\mu \,\mathrm{d}V. \tag{2.45}$$

Now the coefficient matrix C with all its elements $c_{\mu i}$ can be calculated and optimized iteratively. This means to take the coefficient matrix C element by element and optimize it for each of these elements in terms of the energy of the whole system. This will be repeated through the whole matrix until convergence criteria are met.

2.2.3.3 Electron Correlation (EC)

Electron Correlation (EC) is a name for the effect unaccounted for by the approximation of a SCF (see 2.2.3.1). Electrons are not only influenced by the smeared electron charge distribution, but also have explicit interactions with the electron not described by Coulomb repulsion. These interactions can be observed additionally to the average interaction with the charge density. In this context, two types of interactions can be distinguished. The first is the so called *Fermi*-hole, which reduces the probability of finding two same-spin electrons at extremely small distances to zero and reduces the probability significantly in close proximity. The second effect is the *Coulomb*-hole, caused by *Coulomb* repulsion, which reduces the probability of finding two electrons, irrespective of their spin, close to each other, but does not reduce it to zero. These effects are visualized in Figure 2.5.
Additionally, EC describes the fact, that when two configurations of electronic structure are energetically close, only describing one of them is insufficient to describe the system. This is the case as the other configuration will have significant contribution to the electronic structure of the system, since they are easily reachable from the ground state. In DFT this problem can be solved in the term for $V_{XC}[\rho(\vec{R})]$, as it can combine additional potentials to mimic these two effects. In other, so called *Post-Hartree-Fock* methods, either configuration interaction, the coupled cluster formalism or perturbation theory is used to address the problem, which will not be discussed in further detail.

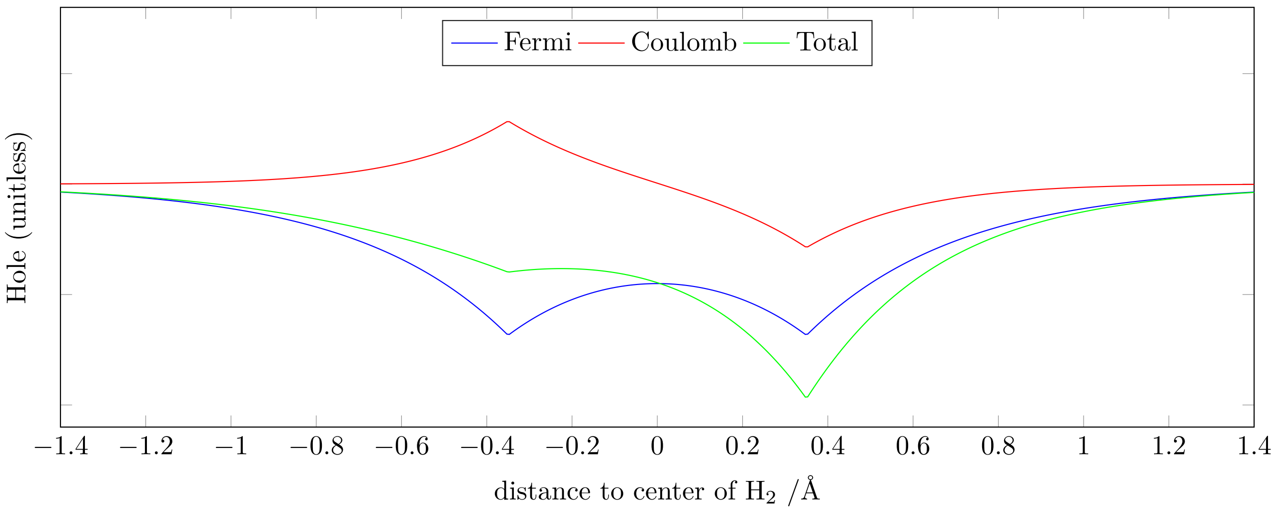


Figure 2.5: Qualitative scheme plotting the magnitude and sign of the Fermi- and Coulomb-hole in the H_2 molecule, if one electron resides at around +0.3 Å from the molecule center. The total hole is given by the sum of both curves and describes the complete correlation effect.

2.2.3.4 Density Functional Theory (DFT)

The foundation of DFT-methods is the First *Hohenberg-Kohn* theorem. [100] The theorem states that for a non-degenerate ground state, the external potential is uniquely mapped to the electron density, and thus the total energy is a functional of the non-degenerate ground-state electron density. This allows the formulation of a functional that describes this potential. The estimation of the molecular potential then is not a problem of $3N$ variables anymore, but only depends on the ED at a certain point in space given by 3 numbers, the coordinates. [100] So the energy of a system then can be expressed as:

$$E(\rho) = \int V(\vec{R})\, \rho(\vec{R})\, \mathrm{d}\vec{R} + F(\rho) \tag{2.46}$$

$F(\rho)$ is the functional description of the difference in energy between a set of orbitals similar to a wavefunction Ψ, only involving the classical Hamiltonian (see equation 2.27), and the energy corresponding to the ED one wants to describe. So if this functional was known or well enough approximated the determination of a a set of orbitals corresponding to a non degenerate ED would be analytically solvable.

The approximation of this functional is performed in DFT-methods, which describe the molecular set of orbitals in a different way than in the *Hartree-Fock* framework: They approximate every one electron orbital ψ_i, the so called *Kohn-Sham*-functions, as a solution of the *Schrödinger* equation using an effective potential v_{eff}, therefore there is no determinant as a solution but discrete one electron orbitals themselves. The resulting equations can be written as:[1]

$$\left(-\frac{1}{2}\,\bigtriangledown^2 + v_{eff}(\vec{R}) - \epsilon_i\right)\psi_i = 0 \tag{2.47}$$

In an N electron system the ED can then be obtained from the occupied orbitals by:

$$\rho(\vec{R}) = \sum_{i=1}^{N_{occ}} \left(\left|\psi_i(\vec{R})\right|^2\right) \tag{2.48}$$

v_{eff} is depending on the ED:

$$v_{eff}(\vec{R}) = V(\vec{R}) + \int \frac{e^2\rho(\vec{R}')}{\left|\vec{R}-\vec{R}'\right|}\mathrm{d}^3R' + V_{XC}[\rho(\vec{R})] \tag{2.49}$$

---

[1] In the following equations the atomic units (a.u.) are used for ease of reading. This system of units accounts for the natural constants and unit conversion factors for atomistic scales and allows focus on the functional relation between properties.

In this context $V(\vec{R})$ denotes the potential due to the nuclei, the second term is the electron-electron interaction and the third term is the exchange and correlation term or potential. This is the major approximation in the framework of DFT, since – excpet the free electron gas system – this term is not known exactly, but can only be approximated. [349] The fact that the potential is depending on the ED and the ED itself on the functions resulting from the equation containing the potential an analytic or direct solution is not possible, but can only be approximated iteratively until convergence is achieved.

Approximations of the Exchange-Correlation Potential

There are several definitions of the Exchange-Correlation potential used for DFT that range from simple approximations based on the homogeneous electron gas to sophisticated mixtures of different levels of theory. They can be categorized by the underlying principle used to derive the potential into the following categories:

- Local Density Approximation (LDA)
- Generalized Gradient Approximation (GGA)
- Hybrid-GGA
- Meta-GGA
- Meta-Hybrid-GGA
- Double-Hybrid-GGA

All functionals are based on the concepts of Local Density Approximation (LDA) and just improve the results by including more information. The common strategy in these approaches is the split of the Exchange-Correlation potential into two contributions:

$$E_{XC} = E_X + E_C. \tag{2.50}$$

The exchange part of this potential is approximated by using the analytically solvable exact electron exchange of the homogeneous electron gas, assuming it can be evaluated for non homogeneous cases at each point independently [350] by

$$E_X^{LDA}(\rho) = -\frac{3}{4}\left(\frac{3}{\pi}\right)^{1/3}\int \rho(\vec{R})^{4/3}\mathrm{d}^3R. \tag{2.51}$$

Local Density Approximation

The correlation energy in the framework of LDA can be described by [351]

$$\epsilon_c(\vec{R}) = a\,ln\left(1+\frac{b}{r_s}+\frac{b}{r_s^2}\right), \tag{2.52}$$

where r_s is the Wigner-Seitz parameter, [352] related to the density by

$$r_s^3 = \frac{3}{4\pi\rho(\vec{R})}. \tag{2.53}$$

ϵ_c subsequently needs to be integrated to yield the correlation energy:

$$E_C^{LDA} = \int \rho(\vec{R})\epsilon_c(\vec{R})\mathrm{d}^3R \tag{2.54}$$

Various functionals have emerged that interpolate the obtained accurate values of correlation energy from quantum Monte Carlo simulations at different density magnitudes [353] using different analytical formulations of a and b for equation 2.52, among them:

- Vosko-Wilk-Nusair (VWN) [354]
- Perdew-Zunger (PZ81) [355]
- Cole-Perdew (CP) [356]
- Perdew-Wang (PW92) [357]

Local Spin Density Approximation

In addition to the LDA, a variation was introduced, that takes into account the spin of electrons and depends on the density of spin-up and spin-down electron density. However the same type of functionals can be used, if adjusted for the possibility of different spin densities by introducing two exchange and correlation terms respectively and weighting them using the local spin polarization ζ: [358]

$$\zeta(\vec{R}) = \frac{\rho_\alpha(\vec{R}) - \rho_\beta(\vec{R})}{\rho_\alpha(\vec{R}) + \rho_\beta(\vec{R})}. \tag{2.55}$$

Generalized Gradient Approximation

These functionals do not base the evaluation of the exchange and correlation terms only on the density, but also incorporate the gradient of the density. These functionals also usually are based on the spin sensitive density approach, so the two parts become functionals of four parameters: $\rho_\alpha, \rho_\beta, \bigtriangledown(\rho_\alpha)$ and $\bigtriangledown(\rho_\alpha)$. In most cases the gradient is introduced into the correlation energy by using a version of the reduced density gradient s, that scales the scalar length of the gradient vector by a factor a to match the current context and the density itself:

$$s(\vec{R}) = \frac{|\vec{\bigtriangledown}(\rho(\vec{R}))|}{a\ \rho(\vec{R})^{4/3}} \tag{2.56}$$

This is a fundamental concept of Generalized Gradient Approximation (GGA) functionals. Its implementation resulted in various functionals, where two selected ones that are used in this thesis will be mentioned explicitly:

- Perdew-Burke-Ernzerhof (PBE) [359]
- Becke Exchange and Lee-Yang-Parr Correlation (BLYP) [360, 361]

Hybrid-GGA

In Hybrid-GGA functionals the Hartree Fock exchange energy is used in addition to the density functional exchange energy to achieve better atomization energies, bond lengths and vibrational frequencies. [362, 363]. The exact HF exchange integral can be expressed as:

$$E_X^{HF} = -\frac{1}{2}\sum_{i,j}\iint \psi_i^*(\vec{R}_1)\psi_j^*(\vec{R}_2)\frac{1}{\vec{R}_{1,2}}\psi_j(\vec{R}_1)\psi_i(\vec{R}_2)\ \mathrm{d}^3R_1\ \mathrm{d}^3R_2 \tag{2.57}$$

This energy is mixed in a linear combination with other functionals for exchange correlation energies in the different functionals, two of which will be highlighted, since they are used in the thesis:
B3LYP [364]:

$$E_{xc}^{B3LYP} = E_X^{LDA} + a_0(E_X^{HF} - E_X^{LDA}) + a_x(E_X^{GGA} - E_X^{LDA}) + E_c^{LDA} + a_c(E_c^{GGA} - E_c^{LDA}), \tag{2.58}$$

where a_0, a_x and a_c are the three parameters yielding the "3" in the functional name, which are defined to be 0.2, 0.72 and 0.81, where the GGA terms use the normal BLYP GGA functional [360, 361] and LDA terms refer to

the VWN-LDA model. [354, 364]
PBE0:

$$E_{xc}^{PBE0} = 0.25\, E_X^{HF} + 0.75\, E_X^{PBE} + E_c^{PBE}, \tag{2.59}$$

with the PBE terms referring to the previous definition of the pure PBE-GGA functional. [359, 365]

Meta-GGA

Meta-GGA functionals introduce one more level of sophistication in the description of the exchange correlation potential, as they introduce in addition to the density and gradient contributions the of contributions of the laplacian, the second derivative of the density, into the calculation of potentials. One of the most widely used family of functionals is the Minnesota type of DFT functionals or TPPS. [366–370] But most functionals of the Minnesota family use exact HF exchange, so only M06-L can be considered a pure Meta-GGA functional.

Meta-Hybrid-GGA

Many Meta-GGA functionals have a formulation which incorporates a mixing or substitution of the exchange energy with the HF energy. [366–370] Here the M06 and TPSSH functionals are mentioned with the following HF substitution percentages: M06: 27% HF, M06-2X 54% HF, M06-HF: 100% HF, TPPSH: 10% HF.

Double-Hybrid-GGA

Double-Hybrid-GGA functionals extend the description of the exact HF type exchange term with the use of perturbation theory for the calculation of the correlation energy. [371] The energy then can be expressed as:

$$E_{xc}^{PBE0} = a\, E_X^{HF} + (1-a)\, E_X^{GGA} + b\, E_c^{GGA} + c\, E_c^{PT2}, \tag{2.60}$$

where a, b and c are fitable parameters of the functional where any GGA can in principle be chosen and E_c^{PT2} is the perturbation theory second order term, defined in spin-orbital form by:

$$E_c^{PT2} = \frac{1}{4}\sum_{ia}\sum_{jb}\frac{\left[(ia|jb)-(ib|ja)\right]^2}{\epsilon_i+\epsilon_j-\epsilon_a-\epsilon_b}. \tag{2.61}$$

Development of Functionals and their Application

The different kinds of functionals are intended to describe some feature of a molecule better than their predecessors. In many cases they succeed in doing so, but recently it was shown that the modern functionals deviate more and more from the accurate description of the exact exchange-correlation-potential or electron density of a molecule. [372] Although there is a discussion how significant the deviation really is [373] and results showing that especially the density has bigger errors in more modern functionals, [374] the differences in these functionals are comparably small compared to the errors in experimental charge density and quantum crystallography, so even LDA oder GGA functionals might already be good enough approximations for the use in context of quantum crystallography.

2.2.3.5 Density Fitting / Resolution of Identity (RI)

ORCA and other QM-software provide a series of convenient approximations, that speed up the calculation of wavefunctions enormously. They can be applied to many methods, ranging from HF over pure DFT, hybrid-DFT to post-HF methods like perturbation theory or coupled cluster calculations. The first approximation for evaluation of the two electron integrals in SCF procedures is the so-called Resolution of Identity (RI) or Density fitting. [375–377] This procedure calculates the four index electron repulsion integrals (compare last term in equation 2.44, which in Mulliken notation becomes $(\nu\mu|\kappa\lambda))$ in the SCF procedure as a series of two- and three-center integrals, by applying an approximate resolution of identity for a so called "auxiliary" basis set, that is used in addition to the basis functions of the MOs. [378–380] The central assumption can be summarized in the following three equations written in Mulliken notation taken from reference [380]. P and Q denote atomic auxiliary basis functions, while $\mu, \nu, \kappa, \lambda$ are basis functions of the normal wavefunction LCAO-MO. M_{PQ}^{-1} is the inverse of the coulomb integral matrix of the auxiliary basis functions given by:

$$M_{PQ} = (P|Q) = \int \frac{P\left(\vec{r_1}\right) Q\left(\vec{r_2}\right)}{r_{12}} \mathrm{d}\vec{r_1}\mathrm{d}\vec{r_2}. \tag{2.62}$$

The matrix is found by resolving the identity of the projection matrix

$$\sum_{PQ} |P) M_{PQ}^{-1} (Q| \approx 1. \tag{2.63}$$

This way the four-index electron repulsion integrals of the SCF procedure can be approximated by inserting the found matrix into the integrals, effectively reducing the four-index calculation into a series of three index calculations:

$$(\nu\mu|\kappa\lambda) \approx (\nu\mu|\kappa\lambda)_{RI} = \sum_{PQ} (\nu\mu|P)\, M_{PQ}^{-1}\, (Q|\kappa\lambda)\,. \tag{2.64}$$

This way the evaluation of the time-consuming Coulomb integrals is significantly faster, as the matrix multiplication reduces the complexity of the problem by one order. Since pure DFT functionals with no exact exchange part only need the evaluation of these terms they benefit most from this assumption. The exact exchange integrals needed for HF or post-HF methods, as well as hybrid-DFT functionals, however, remain a complex 4th order problem.

2.2.3.6 Chain of Spheres (COSX)

The evaluation of exact exchange integrals is where the so called Chain of Spheres (COSX) approximation can provide significant speedup. [381] This method, as available in *ORCA*, approximates the exchange integrals in equation 2.65 by a sum over a discrete grid, as seen in equation 2.66:

$$K_{\nu\mu} = \sum_{\kappa\lambda} P_{\kappa\lambda} \iint \mu(\vec{r_1})^* \kappa(\vec{r_2})^* \frac{1}{r_{12}} \nu(\vec{r_1}) \lambda(\vec{r_2})\, \mathrm{d}\vec{r_1}\mathrm{d}\vec{r_2} \tag{2.65}$$

$$K_{\nu\mu} \approx \sum_{g} w_g^{0.5} \mu(\vec{r_g}) \sum_{\lambda} \int \frac{\nu(\vec{r})\lambda(\vec{r})}{|\vec{r} - \vec{r_g}|} \mathrm{d}\vec{r} \sum_{\kappa} P_{\kappa\lambda} w_g^{0.5} \kappa(\vec{r_g}) \tag{2.66}$$

$P_{\kappa\lambda}$ denotes the density matrix associated to basis functions κ, λ and w_g is the weight of an integration grid point with index g. By selecting only important partners in the molecule a cutoff radius for each contracted basis-set shell is determined and grids constructed for evaluation to reduce the error of this approximation. As one can see the complexity of the integral is reduced significantly and the sum over basis functions κ can be calculated separated from integrals of λ. The combination of RI and COSX is called RIJCOSX and very efficient for the calculation of fixed-geometry wavefunctions when using exact exchange integrals. The authors compare performance and timing of these approximations and conclude that geometry optimizations might need careful investigations, but densities and orbitals can be quite accurately calculated with much less computational effort. [382]

2.3 Bonding and Interaction Indicators

Regardless of the origin of a wavefunction or electron density under investigation, tools to describe strength and type of bonding between atoms are crucial to draw connections between the model system and experimentally observed behavior of a substance. The following non-exhaustive list of bonding descriptors was used in the investigations done in this thesis.

2.3.1 Quantum Theory of Atoms in Molecules (QTAIM)

A crucial step towards a deeper understanding of chemical bonding is the definition of the atom within the molecule. QTAIM was proposed by *Bader* as a method solely relying on the shape and behaviour of the ED ρ and its derivatives to partition the space into atomic contributions. [177,178] The description of this technique in this section is based on the theory laid out in references [177] and [178], therefore individual references will be omitted as all information is taken from these two references. By the nature of ρ of a molecule the maximum is located at the atomic positions, decreasing with distance from the nucleus. If there is more than one atom and they are sufficiently close to each other, following $\vec{\bigtriangledown}\rho$ originating at the nucleus yields one line, that is connecting the two atoms. This gradient line is called a bond path in the framework of QTAIM. In this case, there exists a surface between two atoms, that is only intersected with this one particular bond path. All other gradient lines originating from the nuclei diverge along this surface. This arbitrarily shaped surface is called the zero-flux surface. The shape of an atom is defined, so its surface is the zero-flux surface of $\vec{\bigtriangledown}\rho$, described by its normal vector $\vec{n}(\vec{R}_s)$:

$$\vec{\bigtriangledown}\rho(\vec{R}_s) \cdot \vec{n}(\vec{R}_s) = 0, \tag{2.67}$$

for all points with position $\vec{R}_s$ belonging to the surface.
Using this procedure to find and describe the zero-flux surfaces of all atoms the basins that are enclosed by these surfaces can be described. The volume and charge of an atomic basin Ω_j can easily be derived:

$$V_{\Omega_j} = \int_{\Omega_j} \mathrm{d}V \tag{2.68}$$

$$Q_{\Omega_j} = Z - N_j = Z_j - \int_{\Omega_j} \rho(\vec{r})\, \mathrm{d}V \tag{2.69}$$

To classify the different types of bonding between two atoms the properties of so called critical points are of interest. The critical points are extrema in 3D space of ρ. The nuclei of atoms are absolute maxima of ρ in all directions, giving them a curvature of +3 which is called nuclear attractor, when the sign of the values λ_1, λ_2 and λ_3 of diagonalized Hessian $D(\vec{R})$ is added up:

$$\begin{aligned} D(\vec{r}) &= \begin{pmatrix} \frac{\partial^2\rho(\vec{r})}{\partial x^2} & 0 & 0 \\ 0 & \frac{\partial^2\rho(\vec{r})}{\partial y^2} & 0 \\ 0 & 0 & \frac{\partial^2\rho(\vec{r})}{\partial z^2} \end{pmatrix} \\ &= \begin{pmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{pmatrix} \end{aligned} \tag{2.70}$$

The bond path has a point of local minimum and is a maximum in two dimensions perpendicular to the bond path resulting in a curvature of +1. By the very nature of the definition of the zero-flux surface these minima must lie on the surface. A critical point residing in a ring of atoms is a minimum within the plane of this ring, but a maximum perpendicular to the ring-plane, making it a -1 critical point. Finally, there are cage critical points that correspond to a minimum in all three dimensions. They appear in situations like a cuboid of atoms, with the local minimal electron density in its center, but this is not a mathematical requirement as argued by Bader [178] and numerically shown in an actual molecule in reference [383].
The number of critical points of a certain kind n_X within a molecular geometry to be expected can be estimated, since the number of nuclear attractors is usually known from the molecule at hand. The *Poincaré-Hopf* theorem states that

$$n_{NA} - n_{BCP} + n_{RCP} - n_{CCP} = \begin{cases} 0, & \text{in isolated molecules.} \\ 1, & \text{in crystals.} \end{cases} \tag{2.71}$$

If the theorem is not fulfilled, that means there is an insufficient numbers of critical points found by an algorithm, the reason might be that the local extremum is very shallow and missed by the normal stepping of the analysis

software and therefore difficult to detect.
There is a variety of possibilities to analyze the derived atoms and bond paths and their corresponding Bond Critical Point (BCP). Some of these include the analysis of the value of ρ at the BCP, the laplacian $\triangle$ of ρ, defined by

$$\triangle\rho = \lambda_1 + \lambda_2 + \lambda_3, \tag{2.72}$$

where λ_i refers to the i-th diagonal element of the Hessian matrix (see equation 2.70). The Laplacian gives insight into the curvature of the density at the point evaluated, where positive values of $\triangle\rho$ correspond to local depletion of ρ, while negative values correspond to accumulations of ρ. Plotting $\triangle\rho$ along the bond path can give significant insight into the type of bond at hand, but needs to be carefully checked with different bonding descriptors. A different descriptor at the BCP is the ellipticity ϵ, defined as

$$\epsilon = \frac{\lambda_1}{\lambda_2} - 1, \tag{2.73}$$

which becomes zero if the electron density around this bond path has a cylindrical symmetry, while it increases to typical values between zero and one, if the bond is more elliptical.
These descriptors can be calculated from ρ itself. If additional information, e.g. the origin of a certain portion of the electron density towards an atomic basin is to be evaluated the pair density is needed:

$$\rho_2(\vec{r}_1, \vec{r}_2) = \frac{1}{2}\sum_i \sum_j \left(\phi_i^*(\vec{r}_1)\phi_i(\vec{r}_1)\phi_j^*(\vec{r}_2)\phi_j(\vec{r}_2) - \phi_i^*(\vec{r}_1)\phi_i(\vec{r}_2)\phi_j^*(\vec{r}_2)\phi_j(\vec{r}_1)\right), \tag{2.74}$$

as defined by *Bader*. [179] The *Fermi* hole h for an α spin electron can be calculated as

$$h^\alpha(\vec{r}_1, \vec{r}_2) = -\frac{1}{2}\sum_i \sum_j \frac{\phi_i^*(\vec{r}_1)\phi_i(\vec{r}_2)\phi_j^*(\vec{r}_2)\phi_j(\vec{r}_1)}{\rho^\alpha(\vec{r}_1)}. \tag{2.75}$$

Integrating the *Fermi* hole with the corresponding electron density results in the *Fermi* correlation, when the wavefunction is formulated in *Hartree-Fock* single *Slater* determinants, it can be expressed as

$$F^\alpha(\Omega_j, \Omega_j) = \int_{\Omega_j} \mathrm{d}\vec{r}_1 \int_{\Omega_j} \mathrm{d}\vec{r}_2 \, \rho^\alpha(\vec{r}_1) h^\alpha(\vec{r}_1, \vec{r}_2) = -\sum_{a,b} S_{ab}^2(\Omega_j), \tag{2.76}$$

where $S_{ab}(\Omega_j)$ refers to the overlap integrals of the AOs a and b:

$$S_{ab}(\Omega_j) = \int_{\Omega_j} \phi_a^* \phi_b \, \mathrm{d}\vec{r}. \tag{2.77}$$

$F^\alpha(\Omega_j, \Omega_j)$ describes how much of the electron density of an atomic basin is localized within the basin. In a similar fashion the delocalization of electron density from its basin Ω_j into a basin Ω_k can be described by

$$F^\alpha(\Omega_j, \Omega_k) = -\sum_{a,b} S_{ab}(\Omega_j) S_{ab}(\Omega_k). \tag{2.78}$$

It describes the amount of shared ED between atomic basins, which correlates with the strength of covalent bonds. QTAIM calls this the Delocalization Index (DI). [179] In case of ionic bonds, however, it does not increase and other descriptors need to be taken into account.
One major appealing feature of this theory is that atomic properties, due to the derivation and partitioning of the molecule, add up to molecular properties.

2.3.2 Interaction Density

In the framework of this Thesis we use the interaction density in terms of the difference between the electron density distribution in a given environment with respect to the same molecule with identical geometry *in vacuo*. In the literature this term is used only for the difference between an ED of a molecule in the crystal compared to the wavefunction of the same molecule without any outer influence. [289, 384] Within this thesis the term will be used for the difference between two EDs regardless of their origin and, additionally, the term interaction property will be used to quantify the difference between two scalar fields of properties in two different systems.

To quantify these differences in the ED the number of shifted electrons N_e can be calculated as

$$N_e = \int \frac{|\rho_1(\vec{r}) - \rho_2(\vec{r})|}{2}\,\mathrm{d}V. \tag{2.79}$$

The division by a factor of two is due to the fact that between two ED distributions with a constant number of electrons the shift of an electron will appear twice: once ED is removed from an area which is shown as negative interaction density, it will appear in a different place as positive interaction density.

A different approach to quantify the similarity of two property distributions is the Real-Space R-Value (RRS), [385] defined as the ratio between the integral of differences and the corresponding sum of two property distributions, e.g. for the ED:

$$R_{RS} = \frac{\int |\rho_1(\vec{r}) - \rho_2(\vec{r})|\,\mathrm{d}V}{\int |\rho_1(\vec{r}) + \rho_2(\vec{r})|\,\mathrm{d}V} \tag{2.80}$$

For a property with a conserved integral, like the eletron density of a molecule, the R_{RS} is limited to the interval [0,1], being zero for identical distributions and 1 for a complete shift of all values. For a non constant integral property like $\triangle\rho$ (see equation 2.72) or Φ (see equation 2.83) the R_{RS} can take values in the interval [0,∞[, where bigger values correspond to bigger differences in the distribution.

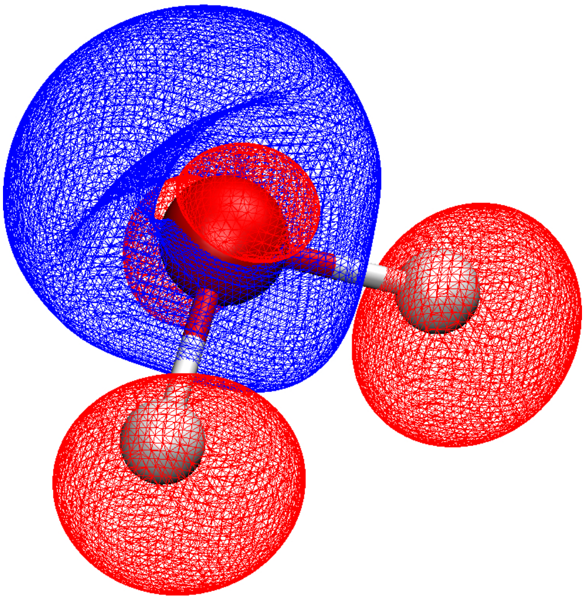


Figure 2.6: Plot of the interaction density isosurfaces at 0.001 a.u. of a water molecule between the wavefunction derived ED of an isolated water molecule and water in an implicit water model. Red areas correspond to higher density in the gas phase, blue areas to higher electron density in the solvation model. Wavefunctions calculated with *ORCA*, volumetric data calculated using *cuQCrT* and visualized with *VMD*.

One can clearly see the effect of the polarizing environment of the solvation onto the distribution of the ED, as depicted in Figure 2.6. The lone pair at oxygen is more pronounced in the solvated state, while the density is withdrawn from hydrogen atoms, as the polarizing environment allow stronger polarization of the bonds, as the resulting dipole can interact with the environment. The R_{RS} for this situation is 0.00523, while the number of shifted electrons N_e is 0.0520 e.

2.3.3 Non-Covalent Interactions (NCI)

To identify interactions between atoms that are not covalently bonded, which can not be described sufficiently by e.g. QTAIM, the reduced density gradient can be used to define descriptors of Non-covalent Interaction (NCI). One descriptor allowing good interpretations is the definition by *Johnson, Contreras-Garcia et al.*, [183, 184] calculating the reduced gradient $s(\vec{r})$ as:

$$s(\vec{r}) = \frac{\left|\vec{\nabla}\rho(\vec{r})\right|}{2(3\pi^2)^{1/3}\rho(\vec{r})^{4/3}}. \tag{2.81}$$

The nature of interaction, that is whether it is repulsive or attractive, is determinable based on the the second value of the Hessian matrix (compare equation 2.70), which corresponds to the classification of critical points within the framework of QTAIM, where the second eigenvalue $\lambda_2 < 0$ for BCPs. Based on this knowledge the sign of λ_2 is used to split the interactions into two sectors, where areas with low $s(\vec{r})$ with sign(λ_2) = + correspond to repulsive interactions, while areas with sign(λ_2) = - describes attractive interactions. This way it is possible to distinguish between different interactions by showing areas of low $s(\vec{r})$ and color code the value of sign(λ_2)ρ onto the surface, which is visualized in Figure 2.7 for an example calculation of a tetramer of water molecules.

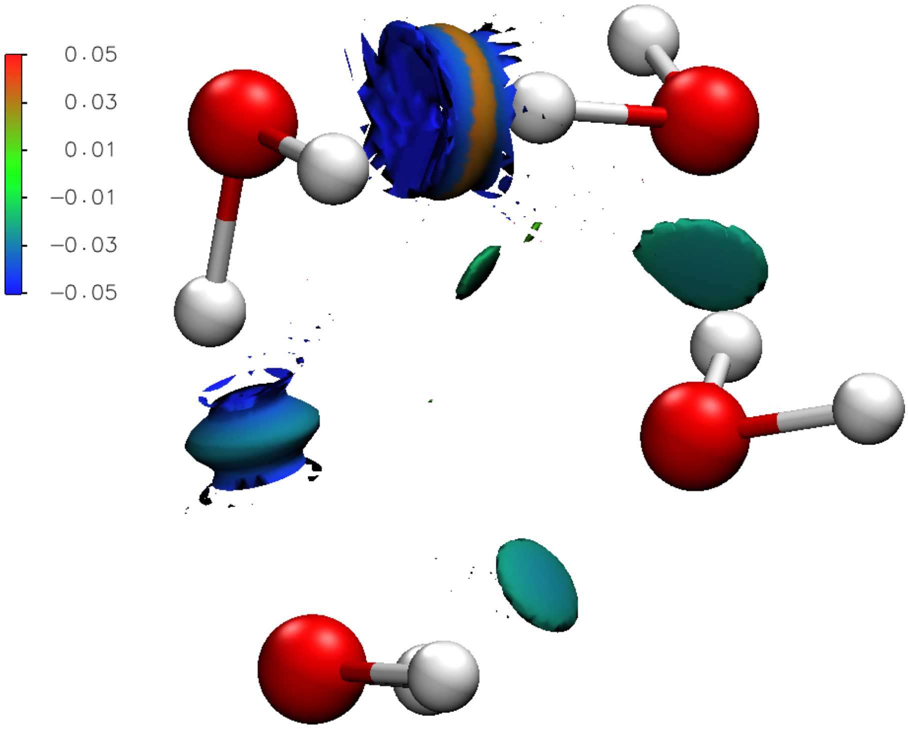


Figure 2.7: Plot of NCI for a example wavefunction of four water molecules with isosurfaces of $s(\vec{r})$ color coded with sign(λ_2)ρ. Isovalue of $s(\vec{r})$: 56 a.u. For color code of sign(λ_2)ρ see color scale bar. Values given in a.u., wavefunctions calculated with *ORCA*, volumetric data calculated using *cuQCrT* and visualized with *VMD*. For software details see appendix, A.10 and A.12.

The calculation of $s(\vec{r})$ can be performed on either the *ab initio* derived density and gradients or on pro-molecular spherical densities using a set of pre defined *Slater*-type functions (equation 2.33), as proposed by the original authors. [386] They show the presence of NCI in both descriptions of the ED, which means the calculation of $s(\vec{r})$ is robust and not unstable or highly sensible in terms of the method or basis set.

The example in Figure 2.7 shows four water molecules interacting in various orientations with each other. The four major interactions will be adressed based on their position in the picture: top, bottom, right and left. The top isosurface shows strong blue regions, which correspond to the hydrogen bond between the hydrogen on the right side and the oxygen on the left. The orange ring around the central part, which has highly positive values of the signed density, corresponds to comparably weak regions of repulsive interaction, which are expected around very strong close interactions, as the density around the hydrogen bond reacts to the strong favourable area within the interaction. [183, 184] The isosurface on the left is a weaker version of the top one, since the distance between the molecules is higher. But again, the blue attractive hydrogen bond is shown with a region of green, less attractive interactions around the central isosurface. the bottom and right isosurfaces show an interaction that is slightly attractive, but far less compared to the directed strong hydrogen bonds on the top and left side. These are rather charge or dispersion dominated interactions without actual strong directed forces. This coincides with the orientation of the hydrogen atoms of the bottom left water molecule facing in a different

direction than the oxygen atom of the right most molecule. These cyan-green shaped isosurfaces usually indicate dispersion or Coulomb interactions, but no strong interaction like hydrogen bonds.
If the NCI is used in order to understand intermolecular interactions rather than intramolecular interactions a useful partitioning of space can be done to determine regions of interest. Two or more entities are defined in the input files of the software (corresponding to molecules or atoms chosen as interacting entities) to separate the complete arrangement presented in a calculation into parts whose interactions are to be analyzed. Those regions where the total density is dominated by one entity, that is 90% of the local density comes from only one of the entities, will be considered intramolecular. Only those regions, which have all contributions of entities smaller than 90% are considered intermolecular. Practically this is done by building partial sums of densities originating from these entities and calculating their ratios. If a voxel is considered intramolecular the value will be set to zero and therefore for practical purposes eliminated.

Thermally Averaged Non-Covalent Interaction (aNCI)

Since the NCI is calculated on a static distribution of electron density it is not possible to distinguish between artefacts of motion induced interactions from stable interactions, e.g. hydrogen bonds throughout a MD simulation. Therefore the averaged Non-covalent Interaction (aNCI) was proposed by *Wu et al.* to account for thermal movement by averaging the NCI over several, individually evaluated geometries. [387] In their work they tested the use of averaged densities or averaged reduced gradients based on individual snapshots of a density, the latter being more useful and used as:

$$\langle s(\vec{r})\rangle = \frac{1}{N_{frames}} \sum_{t=1}^{N_{frames}} s_t(\vec{r}). \tag{2.82}$$

This procedure is able to reduce the significance of thermally introduced interactions, that are not due to persistent molecular interactions but the thermal motion bringing two molecules in close proximity during MD, as it is visualized in Figure 2.8.
It is visible, that the map in case of the aNCI is much clearer and less crowded with low gradient low density interactions, which correspond to close contacts due to thermal motion. These are absent in the aNCI and only strong interactions like the hydrogen bonds formed by the carboxylic acid at sign(λ_2)$\rho \approx -0.05$ in Figure 2.8 are still visible as well as some dispersive interactions around zero, while the picture in the NCI of one single frame is nearly impossible to analyze thoroughly, especially in the region of dispersion interactions around 0. In general $< s(\vec{r}) >$ has higher values than $s(\vec{r})$, as the electron density is fluctuating through thermal motion, which shifts the position of the minimum throughout the simulation, making the minimum less deep, but as shown does not eliminate it, if the interaction is present in many frames of the simulation.

2.3.4 Electrostatic Potential (ESP)

As discussed earlier (see section 1.4.1) the ESP is of crucial importance in the design and efficiency of a drug. The ESP also gives direct hints onto how a bonding situation might have changed comparing two different situations of the same type of functional group, be it in different environments or molecules, as small changes in the distribution of electrons and nuclei have a significant impact on the field expressed by this arrangement. The ESP (Φ) of a molecular wavefunction can be calculated in SI units by a sum of all Coulomb interactions between a positive test particle and the K nuclei with corresponding number of protons Z_K and the total integral over the ED of this wavefunction as

$$\Phi(\vec{r}) = \sum_K \frac{Z_K e^2}{4\pi\epsilon_0 \left|\vec{r} - \vec{r}_K\right|} - \sum_i \int \frac{e^2 \rho(\vec{r}_t)}{4\pi\epsilon_0 \left|\vec{r} - \vec{r}_i\right|}\, \mathrm{d}\vec{r}_t. \tag{2.83}$$

This potential can be used to give insight into electrostatic complementarity when mapped onto a molecular surface (compare Figure 1.9) or can be plotted out in real space to show areas of different polarity. The

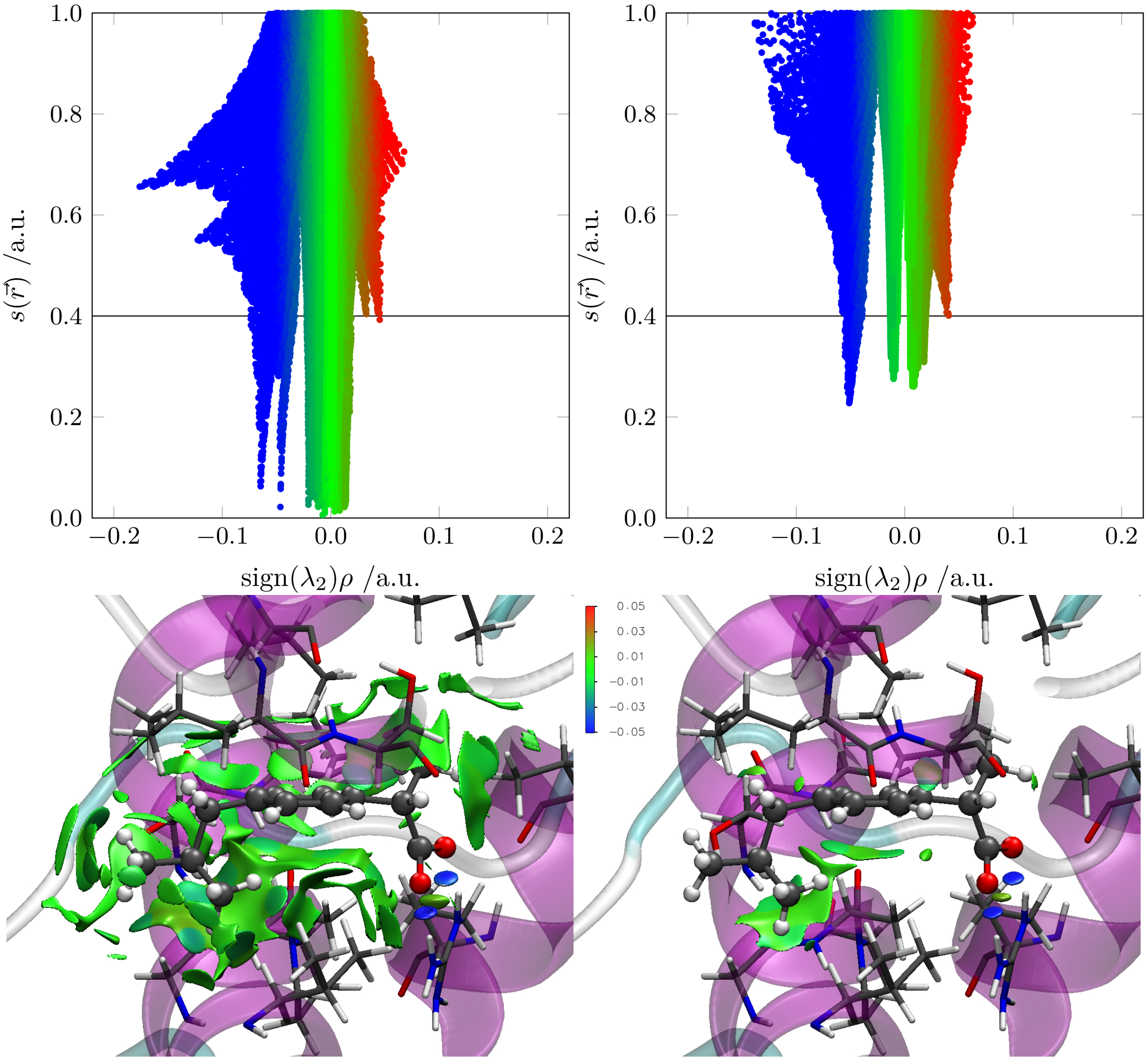


Figure 2.8: Effect of single frame NCI (left) vs. aNCI calculated over 1000 frames of molecular dynamics simulation (right), time steps being 1 ps of ibuprofen inside the active site of COX-II. Top pictures: plots of all volumetric values in the $s(\vec{r})$ vs. sign(λ_2)ρ plane; bottom: 3D representation with isosurfaces of $s(\vec{r})$ color coded with sign(λ_2)ρ. Isovalue of $s(\vec{r})$: 0.4 a u , represented by black line in upper pictures. For color code of sign(λ_2)ρ see color scale bar. Dynamics calculated with *NAMD2*, NCI calculated using *cuQCrT* and visualized with *VMD*. For software details see appendix, A.7 & A.12.

importance of this concept is widely discussed in the literature, for example in reference [388] and the included references. When calculated as a difference between two wavefunctions it can give insight into effects of polarization of the environment or technique used to obtain the wavefunction. This will be called the interaction ESP in the course of this thesis in analogy to the interaction density (compare section 2.3.2).

The similarity or differences of the topology of the electrostatic potential can directly hint towards the similarity or differences of bioisosteric similarity, as shown by the example of carboxylic acid and tetrazole in reference [292]. Two chemically fundamentally different molecules show exceptional similarity in terms of direction and extend of electrostatic features, while a different part of the molecule, namely the methyl group, is much less comparable among the two example molecules shown.

The evaluation of the second term, accounting for the potential experienced by the average field of electrons, is computationally very costly. There are algorithms that can calculate these terms through numerical integration of orbital functions, but the evaluation of these integrals itself is costly. This makes the calculation of Φ for volumetric data and grids extremely time consuming, especially compared to the calculation of electron density and associated properties like Electron Localizability Indicator (ELI) or NCI.

2.3.5 Electron Localizability Indicator (ELI)

A different approach to describe the behavior of electrons described by a wavefunction is based on their localization and localizability. This property is accessible using the curvature of the *Fermi* hole of same spin electrons in general for a single-determinant wavefunction

$$g_\sigma(\vec{r}_j) = \sum_{i<j}^{\sigma} g_{ij}(\vec{r}_j) g_{ij}^*(\vec{r}_i), \tag{2.84}$$

with the curvature g_{ij} being defined as

$$g_{ij}(\vec{r}_j) = \phi_i(\vec{r}_i)\vec{\nabla}\phi_j(\vec{r}_i) - \phi_j(\vec{r}_i)\vec{\nabla}\phi_i(\vec{r}_i). \tag{2.85}$$

In the case of pure HF formalism this reduces to

$$g_\sigma(\vec{r}_j) = \rho_\sigma(\vec{r}_j)\sum_k^{\sigma}\left|\vec{\nabla}\phi_k(\vec{r}_j)\right|^2 - \frac{(\vec{\nabla}\rho_\sigma(\vec{r}_j))^2}{4}. \tag{2.86}$$

The Electron Localizability Indicator (ELI) was defined by *Kohout* [186, 187] to be the electron density necessary to form a certain same spin pair density $D^{\sigma\sigma}$ in a variable region with a volume of

$$V_j \approx \left(\frac{12D^{\sigma\sigma}}{g(\vec{r}_j)}\right)^{3/8}, \tag{2.87}$$

resulting in a charge in this region of

$$Q_j^\sigma = \rho_\sigma(\vec{r}_j)V_j. \tag{2.88}$$

The ELI $\gamma^\sigma(\vec{r_j})$ is the $D^{\sigma\sigma}$ independent part of Q_j:

$$\gamma^\sigma(\vec{r_j}) \approx \rho_\sigma(\vec{r}_j)\left(\frac{12}{g(\vec{r}_j)}\right)^{3/8}, \tag{2.89}$$

which for a HF wavefunction becomes

$$\gamma^\sigma(\vec{r_j}) = \rho_\sigma(\vec{r}_j)\left(\frac{12}{\rho_\sigma(\vec{r}_j)\sum_i^\sigma\left|\vec{\nabla}\phi_k(\vec{r}_j)\right|^2 - \frac{(\nabla\rho_\sigma(\vec{r}_j))^2}{4}}\right)^{3/8}. \tag{2.90}$$

This resembles in some parts the Electron Localization Function (ELF) defined by *Becke* and *Edgecombe* [185]

$$ELF = \left[1 + \left(\frac{\sum_k^\sigma\left|\vec{\nabla}\phi_k(\vec{r}_j)\right|^2 - \frac{(\vec{\nabla}\rho_\sigma(\vec{r}_j))^2}{4\rho_\sigma(\vec{r}_j)}}{\frac{3}{5}(6\pi^2)^{2/3}\rho_\sigma(\vec{r}_j)^{5/3}}\right)^2\right]^{-1}, \tag{2.91}$$

without arbitrarily being restricted to an interval of [0,1].

The topology of the scalar field of the ELI for non correlated wavefunctions is identical to that of the ELF. It can be analysed and yields basins enclosed by zero flux surfaces with critical points, similar to the QTAIM (see section 2.3.1). These basins correspond to individual domains of electron localization, that often coincide with electron pairs, containing values on an absolute scale that can be compared between different compounds, which is not possible with the ELF. [389–391] ELI basins can be sorted into different categories based on their position and size and connectivity to core basins.

The number of connected core-basin related zero-flux surfaces is used to categorize the basins. Core basins are located closely around the nucleus of an atom. The basins around a nucleus are separated into different shells, as the ELI shows a shell structure around atoms of the second period and above. For protic hydrogen atoms, however, due to their single electron being involved in the bond, the core basin is absent and the bond is polarized towards the bonding partner. This changes when looking at hydrides or helium. Monosynaptic basins are basins that are only sharing a face with one other core basin, which is for example the case for lone pairs. Disynaptic basins are connecting two core-basins, which would be the case for covalent bonds between two atoms. In the case of highly polarized bonding the basin might, however, become a monosynaptic basin.

The electron density inside these basins can be integrated and might give raise to hyperconjugative effects and delocalization between different types of bonds. If, for example, electron density in a monosynaptic basin of an element with high valence electron count is missing and a neighbouring basin has an increased electron density integral, a delocalization of lone pair density into the bond might be suggested. These numbers must be interpreted carefully, as in many cases associated lone pair basins never actually reach integer values, so a comparison with model structures and their situation is recommended, as described in the thesis of Dr. *Fugel*. [392]

A different way to interpret the bonding situation based on the topological analysis of ELI and electron density is the Raub-Jansen index. [393] Within this scheme intersections between ELI-D and electron density basins are used to evaluate the contributions of atomic basins of the electron density into bonding and lone-pair basins available from ELI-D. Two contributions are distinguished: The volume of a bond basin contributed by each QTAIM-atom or the integral of the density of this atom within the bond basin compared to the total integral of the bond basin. A significant discrepancy betweeen these two contributions might indicate non classical or polarized bond-character, since the electron density of a bond is mainly attributed to one of the bonding partners.

2.3.6 Interaction Energies

The previous descriptors were based on properties being calculated in real space and resulting in scalar fields. However in many cases not only direction and shape of interaction are of interest but energies that give direct insight into the strength of interactions. A very promising method was the *PIXEL* program that could estimate the interaction between molecules and therefore the sublimation energy of crystal structures based on wavefunctions of the molecule itself without inclusion of the crystal environment. [394] These results can then be read by additional software to visualize the direction and strength of interactions in a visually appealing and intuitive way. [395]

A similar approach to compute energies between pairs of molecules is the concept of model energies for exploration of intermolecular interactions, as proposed by *Turner*, *Grabowsky*, *Jayatilaka* and *Spackman*. [396] In this framework, the total energy between two molecules is constructed by four terms, that are easily accessible by very well studied functions:

$$E_{Tot} = E_{ele} + E_{pol} + E_{dis} + E_{rep}. \tag{2.92}$$

The individual terms are defined as:

E_{ele}: The classical electrostatic interaction energy of the unperturbed charge distributions of both molecules calculated from antisymmetric products of monomer MOs. [397]

E_{pol}: The polarization energy as a sum over all nuclei:

$$E_{pol} = \sum_j -\frac{1}{2}\alpha_{mean,j} \left|\vec{F}(\vec{r}_j)\right|^2 \tag{2.93}$$

where α_{mean} is the isotropic atomic polarizability, e.g. by *Thakkar* and *Lupinetti*, [398] while $\vec{F}$ is the electric field at atom position $\vec{r}_j$.

E_{disp}: Grimme's dispersion correction (in many cases D3(BJ)), summed over all intermolecular atom pairs. [399,400]

E_{rep}: Exchange repulsion energy between unperturbed wavefunctions of the two molecules at hand calculated from antisymmetric products of monomer MOs. [397]

These terms are calculated on the geometry obtained by any technique, be it *ab initio* geometry optimization, molecular mechanics or crystal structure determination, employing molecular wavefunctions for the two molecules under investigation. They are independently calculated and then used to determine the total interaction energy. The wavefunction might be of any level of theory. In their work *Spackman et al.* [396] show the performance of different levels of theory when scaling factors, based on optimization for a benchmark set of compounds with experimentally known interaction energies, are assigned to the different term of equation 2.92 to match experimental values of the total energy. The individual terms are printed without scaling, only the total energy is affected be these factors.

This constructive approach is one way to assess the interaction between two molecules or molecular wavefunctions. The alternative route is the so-called Energy Decomposition Analysis (EDA). Since this technique was only shortly used in the work of this Thesis only a short summary will be given and additional references to further literature given. In this work the *Ziegler-Rauk* approach was used for EDA. [194, 195] This requires a total of three wavefunction calculations: (i) & (ii) for two fragments independently of each other, which corresponds to infinite separation of the two fragments and (iii) where the two fragments are placed in their final arrangement interacting with each other. A full relaxation of the energy through SCF procedures is performed and the changes in energy compared to the sum of the energy terms in the independent fragments evaluated. The changes in energy are partitioned into three major components: The steric repulsion arising from electrostatic and Pauli repulsion terms arising from the interaction between the fragments, the energy orbital interaction between the fragments yields and the dispersive energy of interacting molecules. [401]

2.3.7 Natural Bond Orbitals (NBO)

The theory of Natural Bond Orbitals (NBOs) is based on the localization of MOs into so called Natural Hybrid Orbitals [191] and ultimately into NBOs, [190] that can be used to perform population analysis [402] and analysis on individual bonds, as the localization is chosen to resemble the classical *Lewis* bond picture of bond electron pairs and lone pairs. The theoretical background is quite broad and extensive, so only a short summary is given in this thesis and the interested reader given references for deeper explanations to: [189–191].

There is a certain ambiguity which lewis formula can be used for the analysis, as in polarized cases the covalent and ionic structures might both be inadequate on their own. In these cases the original authors of NBO theory introduced the Natural Resonance Theory (NRT) [403–405] as a procedure to overcome these problems by introducing non-integer bond orders that have both ionic and covalent parts by calculating weights of different NBO representations of Lewis structures and creating the sum of the weighted bond orders to form NRT-bond-orders.

2.3.8 Roby-Gould Bond Index (RGBI)

Based on the idea of a quantum mechanical operator to describe bond orders and quantify interactions directly from the MOs, *Roby* proposed operators to derive these properties from the wavefunction directly without any further approximations or interpretation. [406–409] Based on his work, *Gould et al.* proposed a reformulation of the operator to give ionic and covalent bond orders directly between atoms, atom groups or molecules, depending on the formulation. [410]

The Roby-Gould Bond Index (RGBI) describes a bond in terms of covalent and ionic bond indices, similar to NRT but directly from the wavefunction without a need for localization schemes. These bond indices can be used to define a total bond index. It can be understood as the magnitude of a vector in the 2D-space of covalent in one and ionic bond index in the other dimension, that is spanned for each interaction under investigation. This means the magnitude of the covalent bond index is independent of the magnitude of the ionic bond index, allowing analysis of non-classical bond situations, that would be difficult to describe in Lewis formalism. This works very well for a mixture of covalently or ionically bonded systems Strong non-bonding contributions of non-diatomic systems that in principle give rise to significant energy contributions in the overall wavefunction will lead to a distortion of the bonding picture in this framework as these types are neglected in this formulation.

2.3.9 Calculation of Numerical Data in Parallel

Due to the intrinsic scaling of number of computations needed with the number of voxels of N^3 for numerical volumetric data, excluding overhead, the limits of conventional CPUs are easily reached for desktop computers. A look at the problem at hand, as shown in the simple pseudo-algorithm (in Algorithm 1) of the necessary calculations, however, promises good scaling with the number of processing units assigned to the task, since the loops show no individual dependence on each others result.

Algorithm 1 Pseudocode for calculation of volumetric data, e.g. for electron density from wavefunction.

```
$\psi \leftarrow wavefunction.file$
allocate Grid[$x_{max}$][$y_{max}$][$z_{max}$]
for x=0; x<$x_{max}$ do
    for y=0; y<$y_{max}$ do
        for z=0; z<$z_{max}$ do
            Grid[x][y][z] $\leftarrow$ CALCULATE PROPERTY(x,y,z,$\psi$)
        end for
    end for
end for
function CALCULATE PROPERTY(x,y,z,$\psi$)
    return $|\psi(x,y,z)| * |\psi(x,y,z)|$
end function
```

The triple nested loop assigning values to the grid is ideal for application of the architecture of GPUs. They are built using big arrays of individual processing units, that can be addressed by a mapped array of tasks. *Rubez et al.* have taken the original program *nciplot* of *Contreras-Garcia* and implemented a GPU accelerated version of the calculation of NCI. [188] They reported a 39-fold speed up for a double GPU desktop machine compared to parallel calculation on a CPU. They mapped the problem in a way, so each iteration of two of the three nested loops becomes an individual kernel-call, that is scheduled on the GPU-Multiprocessors. This requires careful management of the memory allocated to store the geometry and grid on the devices and the host memory. Their code could be represented as shown in Algorithm 2.

The pre-defined variables used are A_s and B_s for the Slater type functions (compare equation 2.33) and r being the distance of he gridpoint to the atom i. NumberOfDevices refers to the number of physical GPUs present in the machine performing the calculation and DeviceID being the ID (starting from 0) of a GPU used. Using this algorithm the promolecular calculation of the NCI is possible in fractions of the time it would need on multithreaded CPUs. This is due to several thousand threads running simultaneously on a GPU. The access to values of the grid are exclusive to each thread, so threadsafety is guaranteed.

2.4 Molecular Dynamics (MD)

MD is a procedure to simulate the motion of atoms through time under a given set of conditions by integrating *Newton*'s classical equations of motion using forces obtained by either Molecular Mechanics (MM) or QM. In this way, the simulation of the position and interaction of all particles present in the system can be calculated. MD intrinsically results in a trajectory for all particles in contrast to methods like Monte Carlo, which cannot tell how the motion of one particle to a different position occurs. Due to the equations of motion being difficult to integrate analytically, computer-programs solve them numerically by introducing a time spacing, so called time steps, for which the motion of particles is assumed to be constant. After particles have been propagated one

Algorithm 2 Pseudocode for calculation of NCI on GPUs. <<<x$_{max}$,y$_{max}$ >>> is the notation for mapping of a 2D problem onto the GPU multiprocessors.

```
Molecule ← $geometry.file$
allocate Grid[x$_{max}$][y$_{max}$][z$_{max}$][2] (4th dimension for $\rho$ and RDG)
for all NumberOfDevices do
    LOAD(Promolecular Slater functions)
    ALLOCATE(Device-Grid[x$_{max}$][y$_{max}$][z$_{max}$/NumberOfDevices][2])
    Device-Molecule ← LOAD(Molecule.file)
    KERNEL<<<X$_{max}$,Y$_{max}$ >>>(Device-Grid,Device-Molecule)
    Grid[x][y][z+(z$_{max}$*DeviceID)/NumberOfDevices] ← Device-Grid
end for
function KERNEL<<<X$_{max}$,Y$_{max}$ >>>(Device-Grid,Device-Molecule)
    for z=0;z<z$_{max}$ do
        for i=0;i<NumberAtoms do
            $\rho \leftarrow \rho + \sum_s A_s exp(-B_s r)$
            RDG ← RDG + RDG$_i$ (equation skipped for clarity)
            Hess ← Hess + Hess$_i$ (equation skipped for clarity)
        end for
        SORT HESSIAN(Hess)
        $\lambda_2$ ← Hess[1][1]
        Device-Grid[Kernel$_x$][Kernel$_y$][z][0] ← RDG
        if $\lambda_2 < 0$ then
            Device-Grid[Kernel$_x$][Kernel$_y$][z][1] ← $-\rho$
        else
            Device-Grid[Kernel$_x$][Kernel$_y$][z][1] ← $\rho$
        end if
    end for
end function
```

step in time all forces are re-evaluated. The way these forces are calculated can be separated into two major fields: classical and non-classical force fields.

2.4.1 Classical Force Field

Classical force fields assume, that the interaction between particles can be described by separation of energy into different contributions. There are two major groups of energy contributions, that occur: bonded and non-bonded interactions. Bonded interactions refer to the interactions within a molecule, addressing bonds, angles and dihedrals. Each of these interactions is expressed in a classical spring type of force for bonds and angles and a periodic function for dihedrals.

$$E_{bonds} = \sum_{bonds} K_r (r - r_{eq})^2 \tag{2.94}$$

$$E_{angles} = \sum_{angles} K_\theta (\theta - \theta_{eq})^2 \tag{2.95}$$

$$E_{dihedrals} = \sum_{dihedrals} K_\phi [1 + cos(n\phi - \gamma)] \tag{2.96}$$

K_r, K_θ and K_ϕ are the force constants associated to the type of bond/angle/dihedral, while r_{eq}, θ_{eq} and γ are the equilibrium bond length/angle/dihedral, respectively. n is the multiplicity of the dihedral angle, accommodating the symmetry of groups attached to the atoms which are directly neighboring the dihedral angle. In some force fields, there are also terms called improper angles, which are in fact dihedral angle terms, which have the selection of atoms in a different setting than the dihedral angles, namely a definition of planes through four atoms in a branched arrangement (see Figure 2.9).

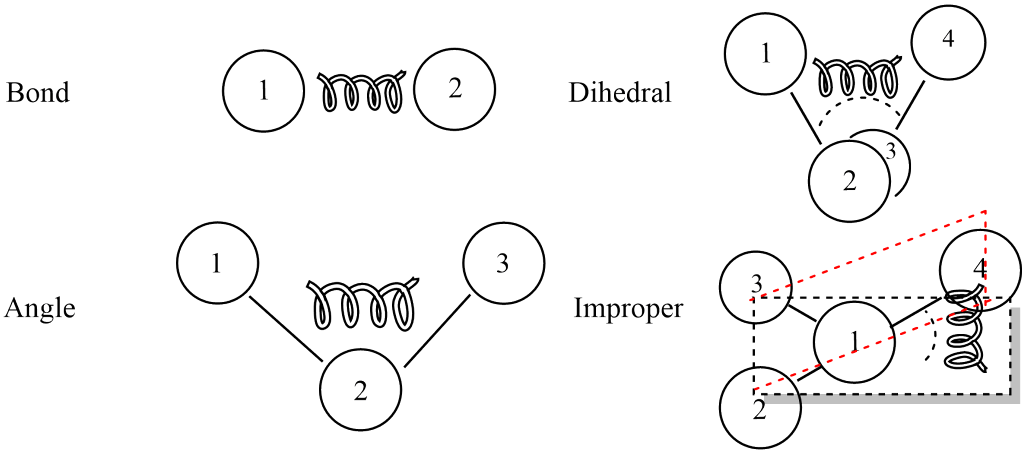


Figure 2.9: Visualization of the parameters of a classical force field as springs between atoms, bonds and planes of atoms.

Additionally, non-bonded terms are defined to consist of two components: electrostatic interactions based on *Coulomb* energy and the dispersive interaction based on *van der Waals* interactions. The electrostatic term is well known and described, while the dispersive term is in most cases assumed to follow a *Lennard-Jones* type potential with exponents 6 and 12.

$$E_{vdW} = \sum_{i<j}\left[\frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^{6}}\right] = \sum_{i<j} 4\epsilon_i \left[\frac{\sigma_i^{12}}{r_{ij}^{12}} - \frac{\sigma_i^{6}}{r_{ij}^{6}}\right] \tag{2.97}$$

$$E_{coulomb} = \sum_{i>j} \frac{q_i q_j}{4\pi\epsilon R_{ij}} \tag{2.98}$$

A_{ij} and B_{ij} or ϵ_i and σ_i characterize the *van der Waals* interaction by using pre-defined terms of the force field for distance and depth of the energetic minimum of one atom type i with respect to all atoms j; q_i and q_j are their respective charges as defined in the force field. ϵ is the electric susceptibility of the medium used and r_{ij} is the distance between the two atoms.

The terms 2.95-2.98 are summed up to give the total potential energy of a particle in a given arrangement by this force field as:

$$E_{pot,total} = E_{bonds} + E_{angles} + E_{dihedrals} + E_{vdW} + E_{coulomb} \tag{2.99}$$

This definition allows pre-definition of a set of force constants, which can be stored in a file and read during calculations. These force constants are the approximated parameter to model the potential energy terms of the system. The only things that need to be calculated during the MD are positions and orientations, which are needed for the integration of *Newton*'s equations anyway. Since both types of non-bonded interactions are in principal infinite with respect to atomic distances there usually is a cut-off criterion defined. At this distance the calculation is either truncated without any correction or linearly decreased in a switch-function to zero potential for *van der Waals* terms. The electrostatic terms can be transformed into a particle mesh Ewald summation [411] in the case of a simulation with Periodic Boundary Condition (PBC). This is the constraint, that all particles in a simulation volume experience all forces, as if the cell was repeated infinitely by translation in each direction. The particle mesh Ewald summation separates the summation of short range energies into a normal sum and completes long range contributions by Fourier transforming the remaining interactions and calculating missing contributions in a truncated Fourier series. [412]

The force $\vec{F}(\vec{R})$ acting on a particle can then be derived from the energy equation 2.99 as its negative gradient.

$$\vec{F}(\vec{R}) = -\vec{\nabla} E_{pot,total}(\vec{R}) \tag{2.100}$$

$E_{total}(\vec{R})$ here represents the energy arising from the applied force field of an atom with the position $\vec{R}$. If $\vec{F}(\vec{R})$ was applied to all particles according to *Newton*'s second law $\vec{F} = m * \vec{a}$, with m corresponding to the mass of the particle, their position could be altered to minimize the overall force on the system and one would obtain the minimum of the force field PES. This minimum would, if the force field was perfectly describing the nature of all atoms and if it was the global and not local minimum, correspond to the minimum structure at 0 K of the simulated structure. This procedure is called MM. However, in MD the focus is not the optimization of structures with respect to energy but the movement of atoms over time. Usually initial MM is performed if the input was taken from different sources or altered by hand. This is required since without ensuring there is no direct clash of atoms, these would lead to incredibly high forces that could break the accuracy of algorithms of integration of the equations of motion. This process is not always necessary but common practice. To simulate the motion of atoms through time a different mechanism to change atom positions is required, that takes into account temperature. The temperature of a system is defined by the *Boltzmann* distribution assigning a probability to find a particle of certain speed in a system of a certain temperature. Following *Newton*'s second law of motion and using the velocities from a *Boltzmann* distribution the position of a particle at a point in time t can be calculated from the previous point in time $t-1$, where their difference is Δt, by:

$$\vec{R}(t) = \vec{R}(t-1) + \vec{v}(t-1)\Delta t + \frac{\vec{a}(t-1)\Delta t^2}{2} \tag{2.101}$$

One method to follow the motion of atoms is the *Verlet* approach. [413] In this approach, the next time step can be calculated using the positions at the current and previous time steps and the force field derived forces.

$$\vec{R}(t+1) = 2\vec{R}(t) - \vec{R}(t-1) + \vec{a}(t)\Delta t^2 \tag{2.102}$$

The positions of all particles are well known during the simulation, so only the evaluation of force field would be time consuming. The first time step needs to be treated slightly differently, since no prior knowledge of atomic positions is possible. Therefore the positions for time step $t_0 + \Delta t$ are generated by assuming a *Boltzman* distribution of $\vec{v}$ and assigning corresponding values to atoms in a random manner. Then a first propagation in time can be done using equation 2.101, with $\vec{a}$ being derived from the force field. In the next iterations the propagation of positions can be done using the *Verlet* algorithm shown in equation 2.102.

For this mechanism to produce reliable results the crucial parameter is a suitable selection of time step size Δt. In general the time step should be capable of resolving all motions of the system in time, corresponding to the time step being smaller than the highest frequency of atomic motion in the system of interest. Since usually hydrogen bonds and interactions of hydrogen atoms with other molecules in a condensed phase are of interest the time step needs to be in the range of 1 fs. As this then easily becomes the limiting factor in terms of possible time scales that can be simulated the need for more robust algorithms arose, that can produce reasonable results with bigger time steps, as this would linearly increase possible simulation timescales. Similar to crystallographic restrains and constrains of hydrogen atoms in riding models the movement of hydrogen atoms can be restricted for the simulation by applying methods like the "shake-algorithm" [414] and bigger time steps, e.g. 2 fs are possible.

In the *Verlet* algorithm the velocities are only explicitly known for the very first time step. This is unfortunate, since the calculation of kinetic energy of the system is necessary to ensure energetic stability of the system in an equilibrium state. In a system under equilibrium conditions the total energy should be conserved:

$$E_{system} = E_{kin} + E_{pot,total} = const. \tag{2.103}$$

Since this is only true if the system is in equilibrium temperature and there would be no integration errors or inaccuracies in the algorithm in all practical cases, one needs to ensure this by additional algorithms acting onto the kinetic energy. Therefore a new set of equations is needed to evaluate the velocities of all particles at minimal computational cost, which can be done using the *Velocity-Verlet* algorithm, [415] as used in this thesis:

$$\vec{R}_i(t+\Delta t) = \vec{R}_i(t) + \vec{v}_i(t)\Delta t + \frac{\vec{a}_i(t)\Delta t^2}{2} \tag{2.104}$$

$$\vec{v}_i(t+\Delta t/2) = \vec{v}_i(t) + \frac{\vec{a}_i(t)\Delta t}{2} \tag{2.105}$$

$$\vec{a}_i(t+\Delta t) = -\frac{\vec{\nabla} E_{total,i}}{m_i} \tag{2.106}$$

$$\vec{v}_i(t+\Delta t) = \vec{v}_i(t+\Delta t/2) + \frac{\vec{a}_i(t)\Delta t}{2} \tag{2.107}$$

Applying this algorithm, the kinetic energy E_{kin} can be calculated from the velocities with $E_{kin} = \frac{1}{2}m_i\,|\vec{v}_i|$ and the potential energy $E_{pot,total}$ can be evaluated from the force field (see equation 2.103).

Ensembles

With this set of equations one can use different approaches to simulate the system according to different ensembles of statistical thermodynamics:

- **NpT** (isotherm isobar)
- **NVT** (canonical)
- **NVE** (microcanonical)

Their names are derived from the constraints applied to the system. **N** is the number of particles and always constant in MD simulations. **p** is the pressure, which can be kept constant by rescaling the size of a simulation box in which particles are allowed to move according to the force they experience over the area of the surface enclosing the simulation box with volume **V**. In systems with periodic boundary conditions the forces acting between particles in the original box at the edges with the particles appearing by the periodic constraint are used to evaluate these forces. In non periodic cases the potential used to keep particles inside the box is directly used for the evaluation of pressure. **E** is the overall energy in the system (E_{System}, see equation 2.103), while T is the temperature of the system, which can be calculated using the velocities:

$$T(t) = \frac{2}{3Nk_B}\sum_{i=1}^{N}\frac{m_i|\vec{v}_i(t)|^2}{2} \tag{2.108}$$

k_B refers to the *Boltzmann* constant, N is the number of particles. To control the pressure of the system, the evaluation of forces acting through the bounding box of the simulation is necessary and if the forces do not correspond to the requested pressure, which is also the forces acting from the box onto the particles, according to *Newton*'s third law of motion, the box size is adjusted accordingly. These mechanisms are called barostats in reference to the real measurement device. Temperature control is achieved by watching the temperature according to equation 2.108 and scaling atomic velocities to match the temperature again. These mechanisms as called thermostats. If **V** is to be kept constant no additional actions need to be taken, leading to variable pressure, which is also true for constant **E**, as the system then is just isolated and can therefore vary in temperature and distribution of energy between E_{kin} and $E_{pot,total}$, while E_{System} is still constant.

Non-classical Force Fields

Applying the classical approach of springs as bonds and angles but using more sophisticated approaches like tailor made QM derived parameters to estimate the associated force constants it is possible to define force fields, for example using tight binding models as in *QMDFF* [234] or *DFTB*. [416] It is also possible, but in most cases not feasible in terms of computational resources, to base the derivation of forces purely on quantum mechanical calculations (see section 2.2), but this is several orders of magnitude slower and more resource demanding. Therefore hybrid methods like Quantum Mechanics Molecule Mechanics hybrid method (QM/MM) have been introduced (see section 2.5) and will be used in this thesis.

2.4.2 Free-Energy-Perturbation Methods (FEP)

In many cases, the energy associated with the whole process of the introduction, recognition and binding of an inhibitor molecule into a protein is necessary to understand its mechanism of action. It is crucial to get insight into the thermodynamic driving forces of the inhibition process at hand. This corresponds to the whole energy of the induced fit process (compare Section 1.4.1). The thermodynamic contributions include the rearrangement of the protein and entropic terms arising from binding or release of molecules and bonds. Especially if the difference in binding energy between two substances is of interest, the terms for rearrangement of the protein might be of importance. One way to obtain this complete energy in a MD framework is the so called Free Energy Perturbation (FEP) approach. [417,418] In this approach the contribution of selected atoms is weighted along a reaction coordinate λ. λ is defined as the ratio of removal or appearance of energetic contribution of a molecule or fragment, by weighting the contribution of these atoms in the evaluation of the energy and subsequent derivation of forces for the complete system arising from the force field. [419,420] In these so called "*alchemical transformation*" setups the model consist of three groups of atoms, (i) all atoms that are independent of the transformation, e.g. the environment and constant parts of a molecule (ii) the atoms associated to the starting situation and (iii) those that are the target state. Their connectivity and force field parameters must be defined and provided. This approach can be used to study thermodynamic effects on the mutation of a protein, as well as introduction of ligands or transformations of a ligand molecule in the active site of a protein. [421] The free energy A change of a system like this can be explicitly stated in case of finite differences

$$\frac{\mathrm{d}A(\lambda)}{\mathrm{d}\lambda} = \frac{A(\lambda + \Delta\lambda) - A(\lambda)}{\Delta\lambda}. \tag{2.109}$$

The formula of the free energy in the canonical ensemble is given by

$$A(N, V, T) = -k_B T \ln \left[h^{-3N} \iint e^{-\frac{E_{System}}{k_B T}} \mathrm{d}\vec{p}\mathrm{d}\vec{r} \right], \tag{2.110}$$

where h is the *Planck* constant, V the volume of the simulated system, N the number of atoms, T the absolute temperature and k_B refers to the Boltzmann constant. E_{System} refers to the energy at each time step, as defined in equation 2.103. The difference between free energies can be expressed by combination of equations 2.109 and 2.110:

$$\frac{\mathrm{d}A(\lambda)}{\mathrm{d}\lambda} = -\frac{k_B T}{\Delta\lambda} \ln \left\langle e^{-\frac{(E_{\lambda+\Delta\lambda} - E_\lambda)}{k_B T}} \right\rangle_\lambda, \tag{2.111}$$

where $< ... >_\lambda$ denotes the average over all ensembles of reference state λ. Therefore the description of both states, starting and target structure, must be a iterative procedure with sampling of the ensembles during the calculation at all intermediate steps. In the framework of FEP the free energy difference between the two states, starting $\lambda_A = 0$ and target $\lambda_B = 1$, can be calculated as a sum over all intermediate steps:

$$\Delta A_{\lambda_A \rightarrow \lambda_B} = -k_B T \sum_{\lambda=\lambda_A}^{\lambda_B - \Delta\lambda} \ln \left\langle e^{-\frac{(E_{\lambda+\Delta\lambda} - E_\lambda)}{k_B T}} \right\rangle_\lambda. \tag{2.112}$$

Obviously this approach is really sensible in terms of stepsize selection for λ and needs full sampling of the ensemble subspace of all intermediate steps. If insufficient states at each step of λ were sampled the system might show artifacts due to sudden strong changes of forces that lead to non-physical behavior. Any change of λ by the nature of the method introduces these artifacts into the system. To exclude these sudden changes of the energy of the system by changing λ at each step first an equilibration during a sufficiently long period of simulation time is performed. This portion of the simulation is not evaluated in the determination of the free

energy according to equation 2.112, only during the following production simulation data is taken into account. This ensures that the evaluation of the free energy is only performed after artifacts through the alchemical transformation were equilibrated and will not influence the result.
Special care needs to be taken when this kind of calculation is done, since the terms arising from intramolecular interactions might introduce significant errors, as this would correspond to additional self interaction of a molecule with its second form, that is atoms of group (ii) interacting with those of group (iii). To account for this, a procedure is introduced that scales contributions of non bonded interactions in the system following the slopes in Figure 2.10. It has to be mentioned, that the selection of for example the value of λ where the coulomb terms meet and also the start and end point of the vdW shifts can be selected while setting up the calculation, but were chosen as depicted in Figure 2.10 for this thesis.

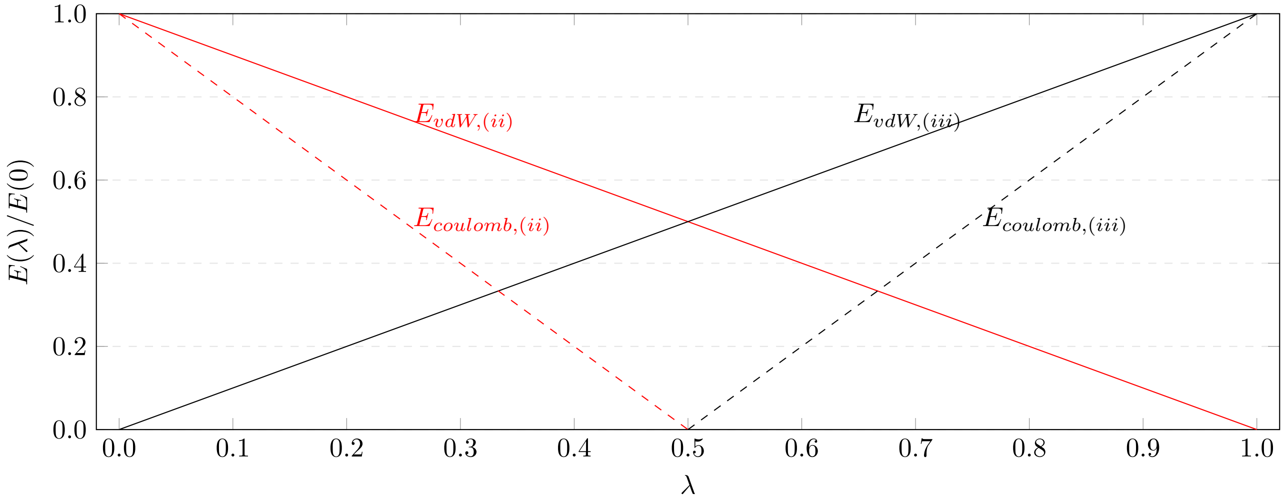


Figure 2.10: Plot of the selected parameters of transition of the bonded and non-bonded energies against λ in FEP MD simulations performed during the work in this thesis.

2.4.3 Root Mean Square Displacement (RMSD)

The Root Mean Square Displacement (RMSD) is a measure for structural similarity of two given structures. By following the development of the RMSD over time of a simulation, one can observe sudden structural changes like a conformational change in a protein or the active site of a protein, given the selected atoms for calculation. It is defined as:

$$RMSD(t) = \sqrt{\frac{1}{N}\sum_{i}^{N}\left(\vec{R}_i(t) - \vec{R}_i(t=0)\right)}. \tag{2.113}$$

For a stable system in equilibrium one would expect a fluctuation around a certain constant value, due to thermal movement. If the RMSD rises over time in a linear way, there might be diffusion or denaturation of the protein, while denaturation usually happens in a less linear way.

2.5 Quantum Mechanics Molecular Mechanics Hybrid Methods (QM/MM)

The first documented formulation of a QM/MM approach was published in 1976. [422] This approach treats a selected region of a big system in a QM way, while the rest is treated by MM. This makes it possible to divide the contribution to the energy of the whole system into three parts:

$$E_{system} = E_{MM} + E_{QM} + E_{QM/MM} \tag{2.114}$$

E_{MM} is given by equation 2.99, while E_{QM} results from the already discussed calculations for QM investigations (see section 2.2). By this separation no interaction between these two systems was introduced. It needs to be introduced into the SCF by point charges of the force field and *vice versa*. These charges polarize the QM system and introduce the environment. The terms for the electrostatic and *van der Waals* interaction are the same as in equations 2.97 & 2.98 with the only difference that i only consists of atoms in the QM region and j only all atoms not in i. The potential energy induced due to the polarization contribution is given by polarizability dependent functions of the form:

$$V_{ind}^{E} = -\frac{1}{2}\sum_{i,j} Q_i \left(\frac{\vec{\mu}_j \cdot \vec{R}_{ij}}{r_{ij}^3}\right) \tag{2.115}$$

$$V_{ind}^{W} = -\sum_{i,k} Q_i \frac{\vec{R}_{ik} \cdot \vec{\mu}_{ik}}{r_{ik}^3} + \sum_{k>k'} \vec{\mu}_k \bigtriangledown \frac{\vec{\mu}_{k'k} \cdot \vec{R}_{k'k}}{r_{k'k}^3} \tag{2.116}$$

With:

$$\vec{\mu}_j = \alpha_j \sum_i \frac{Q_i \vec{R}_{ij}}{r_{ij}^3 \epsilon(\vec{R})} \tag{2.117}$$

$$\vec{\mu}_k = \frac{\mu_0 \coth\left(\frac{C\mu_0\left|\vec{E}_k\right|}{k_B T} - \frac{1}{C\mu_0\left|\vec{E}_k\right|}\right)\vec{E}_k}{\left|\vec{E}_k\right|} \tag{2.118}$$

Here α_j refers to the atomic polarizability, Q_i is the charge of atom i, $\epsilon(\vec{R})$ denotes the reduction of electric fields due to back polarization of the surrounding dipoles at the point $\vec{R}$ given by

$$\epsilon(\vec{R}) = \sqrt{\frac{\sum\left|\vec{E}_j\right|^2}{\sum_j\left|\frac{Q\vec{R}_j}{r_j^3}\right|^2}} \tag{2.119}$$

μ_0 is the permanent dipole of water, k_B refers to the Boltzmann constant, T is the temperature. $\vec{E}$ refers to the local electric field. C describes the resistance to reorientation of a water molecule due to restrictions introduced through interactions, such as hydrogen bonds. These parameters are chosen in reference to water, as the QM/MM approach being introduced for enzymes (V_{ind}^{E}) in a water solution (V_{ind}^{W}).

These modifications did not affect the *Fock*-operator, so the resulting formalism of MOs is unchanged. Therefore a modification of the *Fock*-operator $\hat{F}_i$ is needed. The charges introduced combined with the effects of the induced dipoles would give the modified *Fock*-operator based on Equation 2.44:

$$\begin{aligned}\hat{F}_i = &-\frac{\hbar^2}{2m_e}\Delta_{el} - \frac{Ze^2}{4\pi\epsilon_0 r_{K1}} \\ &+ 2\sum_{i=1}^{n}\int \phi_i^2(2)\frac{e^2}{4\pi\epsilon_0 r_{12}}\mathrm{d}V \\ &- \sum_{i=1}^{n}\int \phi_i(2)\phi_j(2)\frac{e^2}{4\pi\epsilon_0 r_{12}}\mathrm{d}V\frac{\phi_i(1)}{\phi_j(1)} \\ &- \sum_{j,k}\frac{Q_{j/k}}{r_{ij}} + \sum_j \frac{\mu_j r_{ij}}{r_{ij}^3} + \sum_k \frac{\mu_k r_{ik}}{r_{ik}^3}\end{aligned} \tag{2.120}$$

i denotes all QM electrons, j refers to all enzyme atoms and k refers to all water atoms.

With this approach it is possible to calculate a wavefunction for a smaller system in an environment of several thousands of MM-atoms. With the gradient of the energy of this system being available it is also possible to perform MD, but this will be orders of magnitude slower than classical force fields (see section 2.4.1).

Chapter 3

Enabling Quantum Crystallography

3.1 Determination of Accurate Atom Parameters from X-ray Diffraction using Non-Spherical Atomic Form Factors

As discussed in the introduction, the determination of accurate atom positions is crucial for the understanding of molecular properties, electrostatic interactions and directly affects the size and shape of molecular surfaces, that are often used in database screening methods to estimate possible matches of measured protein structures with a newly synthesized molecule. To achieve this, the following chapter will show how the field of QCr has been implemented previously and how it was extended in the context of advanced structural refinement during the course of this thesis via the method and software development of non-spherical atoms in *Olex2* (*NoSpherA2*) (for Publication and details see sections 3.1.2 & 3.1.3) and the Quantum Crystallographic Toolbox (QCrT) (see section 3.3).

3.1.1 Using HAR in *Tonto*

The first and most advanced implementation of HAR is the software *Tonto*. [153] This software is a huge collection of different methods, routines, theories and possibilities, written in an object oriented dialect of *Fortran* designed by Daniel Grimwood and Dylan Jayatilaka. At this date, *Tonto* is mainly maintained by Dylan Jayatilaka and his group and collaborators. *Tonto* can calculate wavefunctions, plot properties, make surfaces for visualization software like *CrystalExplorer* [423] and do bonding analysis, e.g. Roby-Gould Bond indices (RGBI). The implementation of advanced structural refinement in the context of HAR was built to use the wavefunction calculated by *Tonto* itself.

The main route for the most detailed selection of options is the use of an instruction file, which is read by *Tonto* sequentially as the program iterates through the commands. This procedure is somewhat tedious, not user friendly, prone to errors and redundant in many cases. This was improved by the introduction of *HARt*, [424] a terminal interface to *Tonto*, that automatically performs HAR, using a set of pre- and user-selected parameters that can be provided as command-line arguments. This software could then either be called in the terminal by a user or as interfaced into *Olex2* in a dropdown menu, [424] where the settings that are available were mostly accessible through a Graphical User Interface (GUI). *Olex2* would prepare short *Crystallographic Information Files (CIFs)* and a merged and pruned reflection file for HARt, as *Tonto* can only work with this kind of reflection information.

lamaGOET

Since the performance of *Tonto* was limited and the development is done by a small team, the idea to use more widespread QM software in conjunction with the refinement engine of *Tonto* emerged. Recently, a useful addition to read in wavefunctions calculated by external software through the interface *lamaGOET* made refinements possible, that would prove very difficult using *Tonto* alone. The software was used in the course of section 4.4. [425] This interface is a script that provides a GUI to select common settings for HAR and manages files, program calls and output merging in a so-called *.lst* file. This way, the computationally demanding wavefunction calculation can be handled by specialized software, while the partitioning and least squares refinement are done in *Tonto*. The name is an acronym for the author of the script, *Lorraine A. Malaspina*, and the software it was intended to use: ***G**aussian*, ***O**RCA*, ***E**LMO-DB* and ***T**onto*. However, since the transfer of wavefunctions from *ORCA* into a file format readable for *Tonto* without loss of computational accuracy was not achieved, the interface to *ORCA* is not existent yet.

Remaining problems

A major drawback of the implementations using *Tonto* for Hirshfeld-partitioning and refinement is the low speed, and limitations when it comes to crystallographic problems such as disorder, atom occupancies and site symmetry. Such refinements are not possible within the framework of *Tonto*, but disordered structures make up a significant amount of the reported crystal structures (292.4225 structures out of 1.071.011 structures in the CSD, as of 18.06.2020). Additionally, the accuracy of Becke grids used during the calculation of atomic scattering factors is insufficient for the calculation of elements heavier than Kr, as was found during tests for the implementation of *NoSpherA2* (see section 3.1.2). Also there is a very limited availability of basis sets for heavier elements in *Tonto*, as there is a special format of basis set files required to be used in *Tonto*. However, these problems play an important role in many structures of modern crystallography, as disorder is present in many structures of organic or biological materials, partial occupation of different ions is a key property of many inorganic materials and heavy elements are widespread in catalysis and metal-organic material design. Many structures therefore challenge *Tonto* in ways where refinements are impossible or unstable.

3.1.2 Implementation in *Olex2* as *NoSpherA2*

3.1.2.1 The *.tsc* File

The definition of a standardized file format to store the contribution of atoms to the scattering in a crystal, called *.tsc*-file, the **t**abulated **sc**attering factors can be read by *olex2.refine* and used for the least squares procedure. [426] It does not matter technically whether the scattering factors (or form factors) are of spherical nature as in the IAM or of any of the existing schemes of non-spherical form factors such as multipoles, HAR, HAR-ELMO, HAR-QM/ELMO or any other method that could produce atomic scattering factors. This way the GUI, file setup, analysis tools and visualization of results that *Olex2* provides become available for HAR.
The following information needs to be contained in the presented form in the *.tsc* file in order to use the new atomic form factors interface:

TITLE: optional title of the structure
SYMM: 'expanded' or list of symmetries[1]
SCATTERERS: space-separated list of all atoms
{ ANYTHING }: colon must be present
DATA: (denotes the end of the header)

h	k	l	A_1	A_2		A_n
h_1	k_2	l_1	$f_1(h_1,k_1,l_1)$	$f_2(h_1,k_1,l_1)$	$\cdots$	$f_n(h_1,k_1,l_1)$
h_2	k_2	l_2	$f_1(h_2,k_2,l_2)$	$f_2(h_2,k_2,l_2)$	$\cdots$	$f_n(h_2,k_2,l_2)$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
h_n	k_n	l_n	$f_1(h_n,k_n,l_n)$	$f_2(h_n,k_n,l_n)$	$\cdots$	$f_n(h_n,k_n,l_n)$

[1] In either case, all symmetry equivalent miller indices must be present in the DATA section. If a list of symmetry operators, expressed as rotation matrices (e.g.: 1 0 0 0 1 0 0 0 1;-1 0 0 0 1 0 0 0 -1) is provided, then the miller indices must be ordered into corresponding blocks - and each block must have symmetry equivalent indices in the same position in each block and generated by the corresponding matrices. This allows for more efficient calculations during the refinement. Otherwise, if SYMM has the value expanded, the indices can be present in any order.

The header of the [name].tsc file is free-format, as long as it contains the space-separated list of atom names in the ‘SCATTERERS:’ line and finishes with ‘DATA:’. Any identifier must be followed by a colon. The identifiers may start with a space. The values $f_1(\mathrm{h}_1,\mathrm{k}_1,\mathrm{l}_1)$ correspond to the complex scattering factor, written as "Re,Im" and h,k,l are the integer miller indices of the corresponding reflection this scattering factor was calculated for. This file format definition is the result of a collaborative effort between the Olexsys Ltd., University of Durham and University of Bern, namely Oleg Dolomanov, Norbert Peyerimhoff, Laura Midgley, Horst Puschmann and Florian Kleemiss. The calculation of $f_1(\mathrm{h}_1,\mathrm{k}_1,\mathrm{l}_1)$ must happen relative to the atomic nucleus position, that is the real space vector in the Fourier transform *must* be relative to the atom position of the atom for which the scattering factor is calculated and the calculation must be repeated for all symmetry occurances of each atom in the unit cell, which can be achieved by calculating the same set of reflections after applying the corresponding symmetry matrix to the miller indices.

3.1.2.2 Principle of HAR in NoSpherA2

The interplay between software and the filetypes, which are needed to interface them, is visualized in Figure 3.1. A more detailed version will be discussed in Figure 3.2.

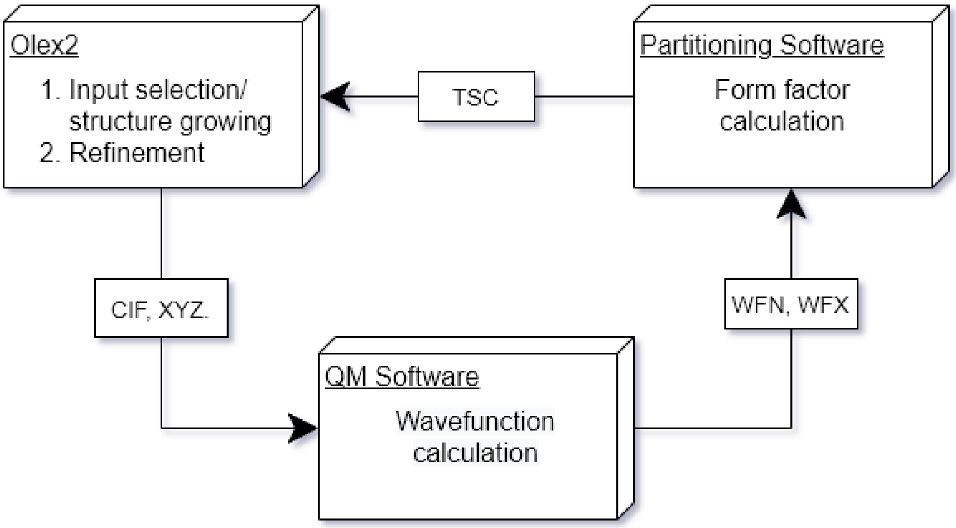


Figure 3.1: Scheme of the interfaces and file-formats handed between software necessary to perform HAR in NoSpherA2. If any other kind of method is to be used QM Software and/or partitioning software can exchanged as needed. CIF/XYZ or even a PDB and the TSC are the only fixed requirements for the interface, all other bits are intentionally designed to be replaceable.

The refinement is initiated from the GUI and a pair of *.cif* and *.xyz* files of the selected entity which is shown on the screen is created in a sub-folder of the current refinement folder. Additionally, an input file is created for the QM-software, as selected in the GUI and stored in the same folder. This input file contains the basis functions, geometry, level of theory and additional information selected by the user. Then a subprocess is started, running the wavefunction calculation using the created input file in any of the possible software: ORCA, Tonto, Gaussian, pySCF and in the future maybe additional code. The results are exported in the *.wfn* or *.wfx* file format for wavefunctions and given, together with the *.cif* previously created, to the partitioning software.

In the first implementation this software was *Tonto*, which was quickly exchanged once the inaccuracy of Becke grids emerged. The new software, called *NoSpherA2*, handles wavefunction file reading, Hirshfeld partitioning itself based on Thakkar [427,428] spherical densities and output of the *.tsc* file. The main difference between partitioning in *Tonto* and *NoSpherA2* is the accuracy of *Becke Grids* and the use of Gaussian-type basis functions for the calculation of spherical densities in *Tonto*, while *Slater*-type functions of the *Thakkar* basis [427,428] are used in *NoSpherA2*. The implementation of Hirshfeld Atom Partitioning in *NoSpherA2* was done in a parallel manner, where OpenMP parallel loops are used and require no external parallelism handling like MPI or external software. Also, a standalone implementation of *Becke Grids* from the git repository *DFTlibs*, called *numgrid*

was used, that is basically a wrapper for the sphere Lebedev Rules. [429, 430]
When running test calculations an error in the electron integration of *Tonto* was found, where a calculation of an Os atom in the gas phase resulted in aprox. 66 electrons instead of 76 electrons that a neutral Os atom should have. A suspected error is the accuracy of *Becke Grids* in *Tonto*. The integration of *Non-spherical atoms in Olex2 (NoSpherA2)* did not show this problem and can integrate the electron density of these heavier elements accurately. To make calculations feasible and efficient, the RI [378–380] and COSX [381] methods are used wherever possible in *ORCA*, *Gaussian* and *pySCF*.

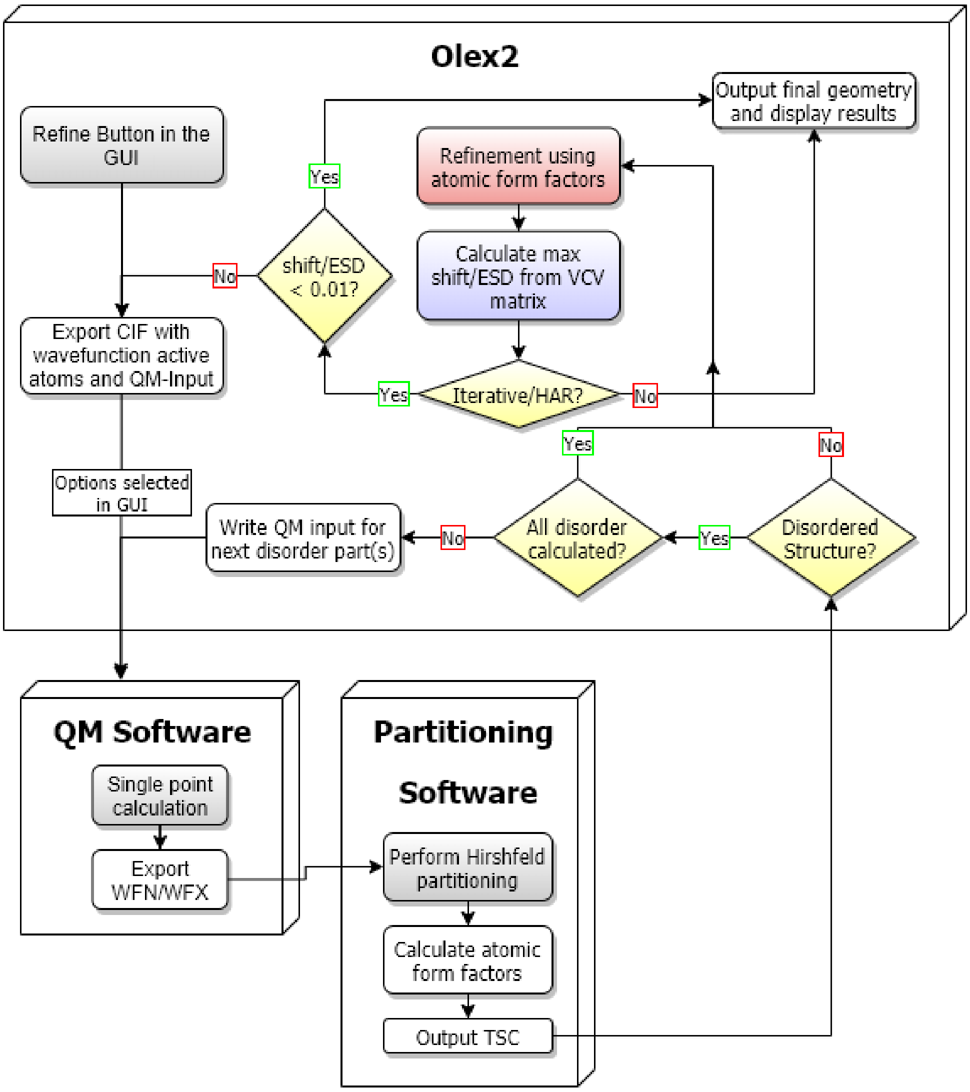


Figure 3.2: Scheme of the checks, processes and workflow of HAR in NoSpherA2.

3.1.2.3 Validation of NoSpherA2

For a validation of the method, an implementation using partitioning done by *Tonto* and also wavefunction calculation performed by *Tonto* was used to make sure identical results were obtained when using *olex2.refine* compared to *Tonto*. Moreover, structural parameters such as bond lengths and ADPs, as well as refinement indicators in terms of R-statistics and residual densities were compared. The benchmarks, as shown in Appendix B.1, showed no difference between refinements of the small organic molecules L-alanine and glycil-L-alanine using the classical implementation in *Tonto* or *Tonto*-calculated *tsc*-files in *olex2.refine*. Bond distances, ADPs and refinement statistics could be reproduced when using the same set of reflections as input for the refinement, which had to be performed manually, since *Tonto* and *olex2.refine* have different pruning and cutoff criteria. By choosing a manual cutoff of $I > 2.1\sigma(I)$ the used number of reflections during the refinement was identical. No difference in the refinement results were observed when wavefunction calculations in *ORCA* were performed and partitioned in *Tonto* for subsequent refinement in *olex2.refine*. A final benchmark involved the entirely *Tonto*-free version of HAR: calculation of wavefunctions by *ORCA*, partitioning by the newly written code *NoSpherA2* and refinement using *olex2.refine*. Here minor differences were observed, but far below any uncertainty of reflections.

Numerical differences are to be expected since the refinement engine in *Tonto* refines against the data based in F, while *olex2.refine* refines against the more reasonable scale of F^2, since this is proportional to the experimentally available intensity of reflections. A different source for minor differences might be the wrong number of electrons based on *Tonto* wavefunctions. This error is not as severe in lighter elements as it is for the fifth and sixth period elements, but might also be a reason for minor differences in results. An additional benchmark involved the use of new features, that only *NoSpherA2* can provide, where an updated weighting scheme of type $w = \left(\sigma^2(F_o) + (aP)^2 + bP\right)^{-1}$ with $P = (F_o^2 + 2F_c^2)/3$ was used. Here the differences seem to be slightly favorable over the normal refinement using $w = \sigma(F_o)^{-2}$, as implemented in *Tonto*.

3.1.2.4 Relativistic Treatment

Additionally, *NoSpherA2* provides the basis sets of the x2c- and jorge-family besides the ones shipped with HARt (cc-p and def2 basis sets, compare 2.2.2), that are optimized for the use with two-component relativistic Hamiltonians and are defined as all-electron basis sets for element up to Z=86 (Rn). Previously available basis sets in *Tonto* with sophisticated accuracy were only available up to Z=36 (Kr). Higher elements were available only by the use of Effective Core Potentials (ECPs), which describe the core of heavier atoms by effective potentials in the wavefunction calculation, reducing computational cost significantly, since the number of atomic orbitals is significantly reduced. However, these ECP can not be used to describe the diffraction of an atom directly, since a major part of its electron density would not be described by the wavefunction. This made investigations on many transition metal structures and also heavier main group elements, that are very interesting in many chemical questions ranging from hydride structures over catalytic complexes to material design, impossible using HAR.

The new basis sets need the use of relativistic Hamiltonians. These can easily become very time consuming when using sophisticated implementations. The Douglas Kroll Hess (DKH) two-component Hamiltonian is a good compromise between accuracy and computational cost and widely available in many quantum mechanical software. [431–433] Using the combination of DKH Hamiltonian and x2c or jorge basis sets, the routine calculation of all-electron wavefunctions for elements of the 5th and 6th period becomes possible. This setup is in principle also available inside *Tonto* if the basis sets are provided to the software and DKH formalism is selected. But the error in the density evaluation mentioned in subsection 3.1.2.2 makes it unreliable to use *Tonto* for these investigations.

When treating metal systems the question of open-shell system becomes imminent. Therefore *NoSpherA2* natively supports the calculation of open-shell wavefunctions using unrestricted formulations of the HF and DFT algorithms in the corresponding wavefunction calculation code. The resulting wavefunction file then contains singly occupied molecular orbitals that can be evaluated in the same manner as in closed shell systems. This approach would in principle also be applicable to multiple determinant methods and all post-HF methods, which have so far not been implemented.

3.1.2.5 Disorder

A major limitation of HAR was the restriction to molecular non-disordered crystal structures, while many structures of interest show a lot of disorder, ranging from ion exchanges, partial occupations of atom positions to mutations and multiple conformations in proteins or side chains of a molecule. These structures could only be treated by HAR when the disordered part was neglected or treated manually by removing contributions from the measured intensities, as performed in the first implementation of the HAR-ELMO method, shown in section 4.4. [134] In NoSpherA2, the refinement of disordered structures is possible through the definition of different parts, that will be treated as individual wavefunction calculations. This is straight-forward for, e.g., a mutation of a single amino acid in a protein, where the situation might be understood as two different versions of the protein being present in the crystal lattice in different positions of the crystal. This can be understood as occupational disorder. Also different conformers of identical molecules can be understood as different entities within one crystal, which need to be described independently. These would correspond to the textbook notion of conformational disorder.

In the situation of multiple mutations or disordered parts of a molecule, the situation might not be that straight-forward. The calculation of one wavefunction using each part might not be sufficient any more. Different parts of the molecule, as defined in the refinement setup mentioned in Figure 3.2, might interact differently with other of these parts, that also show disorder. One example is hydrogen bonds being formed between two residues of a protein in one form/mutation, while these bonds might not be present in the other conformation/mutation. This could never be accounted for correctly by only having one type of form factor for one type of atom. Therefore a splitting of atoms, that are not directly disordered but neighboring the disordered parts is needed for different situations. The definition of groups of local parts that are substitutions for each other with respective weights that sum up to 1 are required. Subsequent permutation of all combinations between parts of different groups ensures reasonable entities for the calculation of wavefunctions, where no covalent bonds are broken.

The idea is basically to define sets/groups of parts where each member of this set is once calculated with each other possible permutation of members of all different sets in order to calculate all possible situations of wavefunctions and understand their effect. However, it may be rationalized that a first implementation of disorder treatment only merging files of different parts would be sufficient, since the effect on the values in the *tsc*-file is getting less and less important, the further away from a disordered description a atom is in the wavefunction. This behavior will be shown in the next section. The structure of the molecule THPP consisting of two disorder parts within the asymmetric unit refined with selected atom labels is shown in Figure 3.3.

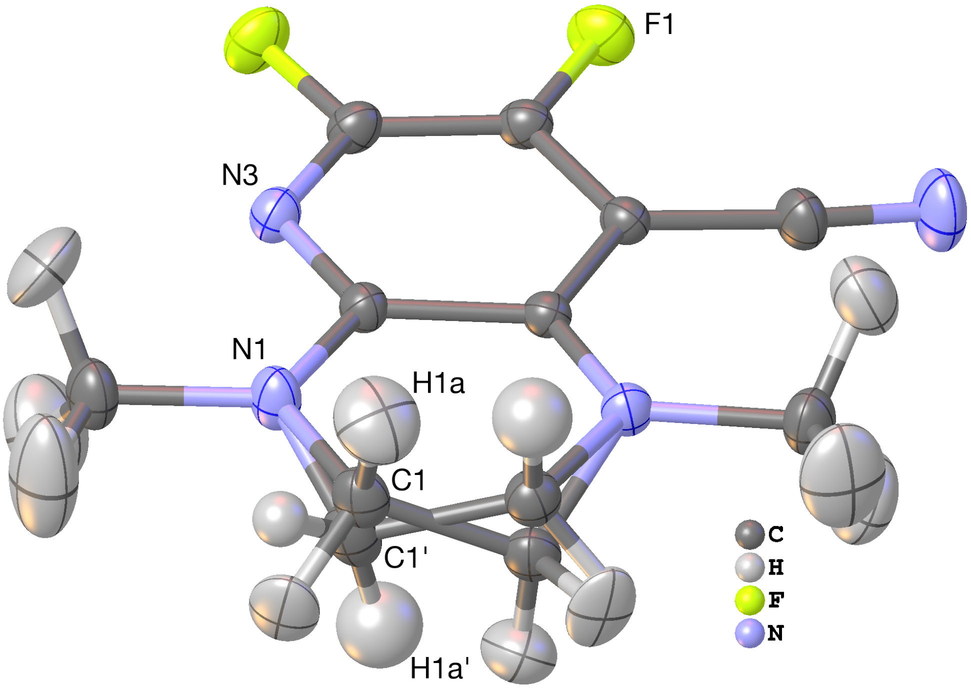


Figure 3.3: Structure of THPP, one of the example structures shipped with Olex2 that was used during the work in section 3.1.3, after HAR using NoSpherA2 with labels of atoms chosen for Figure 3.4. Visualization done using *Olex2* showing thermal ellipsoids for all atoms, except hydrogen atoms of disorder part 2, at 80% probability level.

Size of the Effect of Disorder on Form Factors

Based on the THPP example two parameters were chosen to illustrate the significance of differences in the form factors of disordered atoms and the effect on other atoms. The first parameter is the length of the difference vector between two form factors of atoms in the complex plane with respect to the average of lengths of both individual form factors. This value gets bigger the higher differences are in terms of the absolute length, which corresponds to the charge of the atom contributing to the selected reflection, but also the bigger the difference of the angle in the complex plane with respect to the axis system. This is why a second parameter, the angle between the two vectors, is presented. A plot of these difference indicators is shown in Figure 3.4.
The differences of form factors of the explicitly disordered atoms C1 and H1a are dominant across the whole range of resolution of the reflections. This is true both in relative difference length as in angle differences within the complex plane. The differences of direct neighbors to disordered atoms (N1) have a significantly smaller extend of approximately one order of magnitude, that is, however, not negligible. An approximately exponential decay of importance in the difference of description using non-spherical scattering factors is observed with increasing distance, which coincides with the exponential decay of electron density of atoms. So the influence onto the atomic form factor of disordered atoms is expected to decay significantly over distance, as long as a variation of the bonding motif is not explicitly carried across long conjugated π-systems or other strong interactions, which might propagate differences in description of disordered parts further. In this case the difference of N3 and F1 are at least two orders of magnitude smaller than any difference in the actually disordered part, as there is no electron-conjugation with the disordered methylene groups.

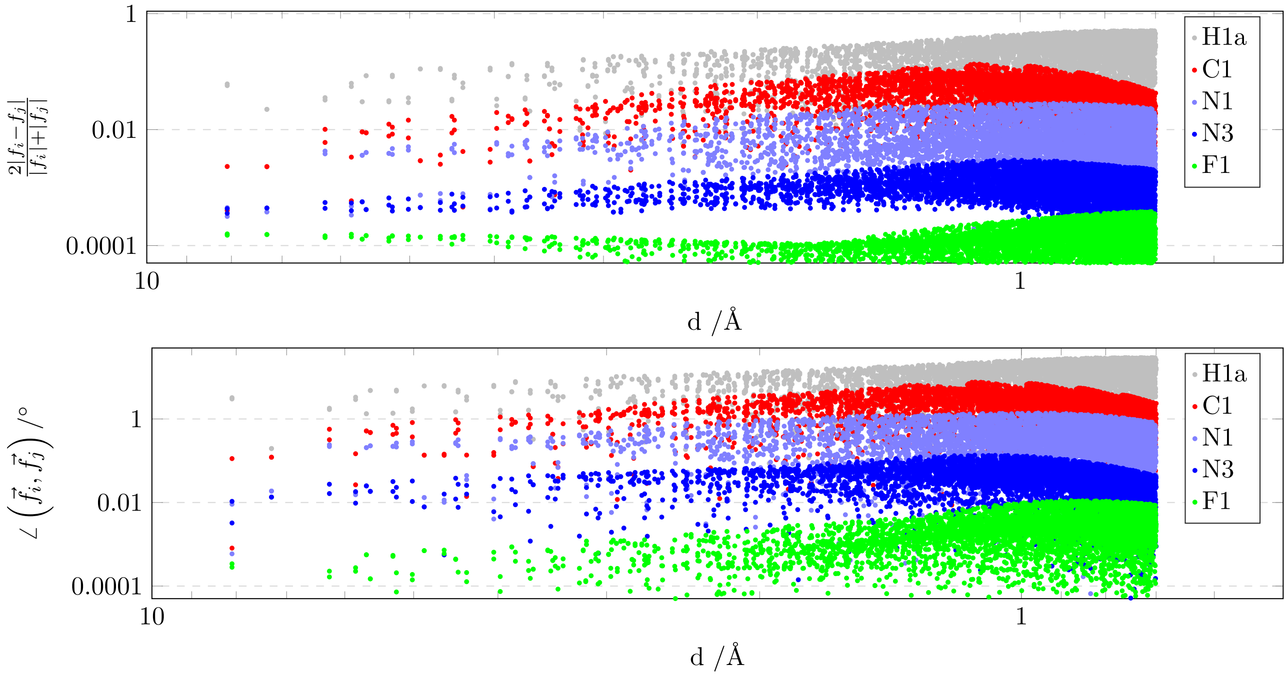


Figure 3.4: Plot of the absolute difference of modulus (top) and angle (bottom) of selected atoms. Values correspond to the differences in the complex plane for non-spherical atomic scattering factors f_i of atom i in the *tsc* file and the disordered equivalent atom j and its scattering factor f_j. These values were exemplarily plotted for the structure THPP (see Figure 3.3) against the d spacing (d_{min}=0.7, d_{max}=8.1) of the corresponding Miller indices.

In principle, a perfect model would include the refinement of two complete molecules, which have their own set of scattering factors. However, this model would easily become too flexible and unstable during the refinement. Therefore, a good practice might be to generate a copy of the atoms directly neighbouring the disordered parts and adding EXYZ and EADP constraints to them, so they can accommodate both scattering factors, while not being independently refined and therefore not significantly decreasing stability of refinements. Unfortunately, until now this has to be prepared by the user manually.

3.1.2.6 Choosing a Level of Theory for HAR

When using HAR, the question of the choice of methods, basis sets and accuracy of integration grids becomes crucial in many ways: (i) The quantum mechanical calculations scale exponentially with the number of basis functions, as there are 3- and 4-index integrals in the SCF procedures. (ii) The correct description of the density is dependent on the description of e.g. the cusp at the atomic position which can only be described accurately when using linear combinations of many basis functions. (iii) The basis set coefficients dictate the radial grid density for the form factor calculation, bigger basis sets automatically require more density points to be evaluated. (iv) The radial accuracy of the integration grid used during HAR as well as the number of spherical points for each radial point significantly increases the accuracy of the density evaluation, however at a huge cost during the Fourier transformation of the density, as each gridpoint must be evaluated for all symmetry generated Miller indices. (v) Higher methods like hybrid-DFT require much more computational time for the wavefunction, while e.g. GGA functionals give less accurate energies in quantum chemical calculations. [372] However, the main concern of HAR is the description of the density in the most accurate way, therefore it will be discussed in Section 3.1.3 that a GGA functional like PBE seems sufficient even for big molecules with heavy elements. Whether it might even be reasonable to use LDA functionals remains to be evaluated and investigated (compare also section 2.2.3.4).

Basis Set and Grid Accuracy Dependence

Since (v) is already discussed in the publication in section 3.1.3, the remaining points will be considered in the following case study. The test data-sets of L-alanine ($C_3NO_2H_7$) at 23 K [434] and the ammonium borate ($(NH_4)_2B_6H_6$, see structures in Figure 3.5) from the publication in section 3.1.3 were refined using PBE and *ORCA* with various combinations of additional parameters. The resulting calculated and measured intensities were used to calculate the weighted RMSD, assuming that the calculated structure factor is without uncertainty and all uncertainty originates from the observed structure factors. The wRMSD results from:

$$wRMSD = \sqrt{\frac{\sum_n \left(\frac{\left(F_c^2 - F_o^2\right)^2}{\sigma(F_o^2)^2} \right)}{n}}. \tag{3.1}$$

Figure 3.5: Structure of L-alanine (left) and ammonium borate (right) after HAR in NoSpherA2. Visualization done using *Olex2* showing thermal ellispoids at 80% probability level.

n denotes the number of unique reflections in the *.fcf* file after the refinement converged. Perfect agreement of all reflections would result in an wRMSD of 0, statistical agreement within the experimental uncertainties should yield a wRMSD of approx. 1. A series of basis sets was used for the benchmark using Normal and High grid accuracy from the *Olex2* GUI for each basis set. This selection will be applied automatically to both the

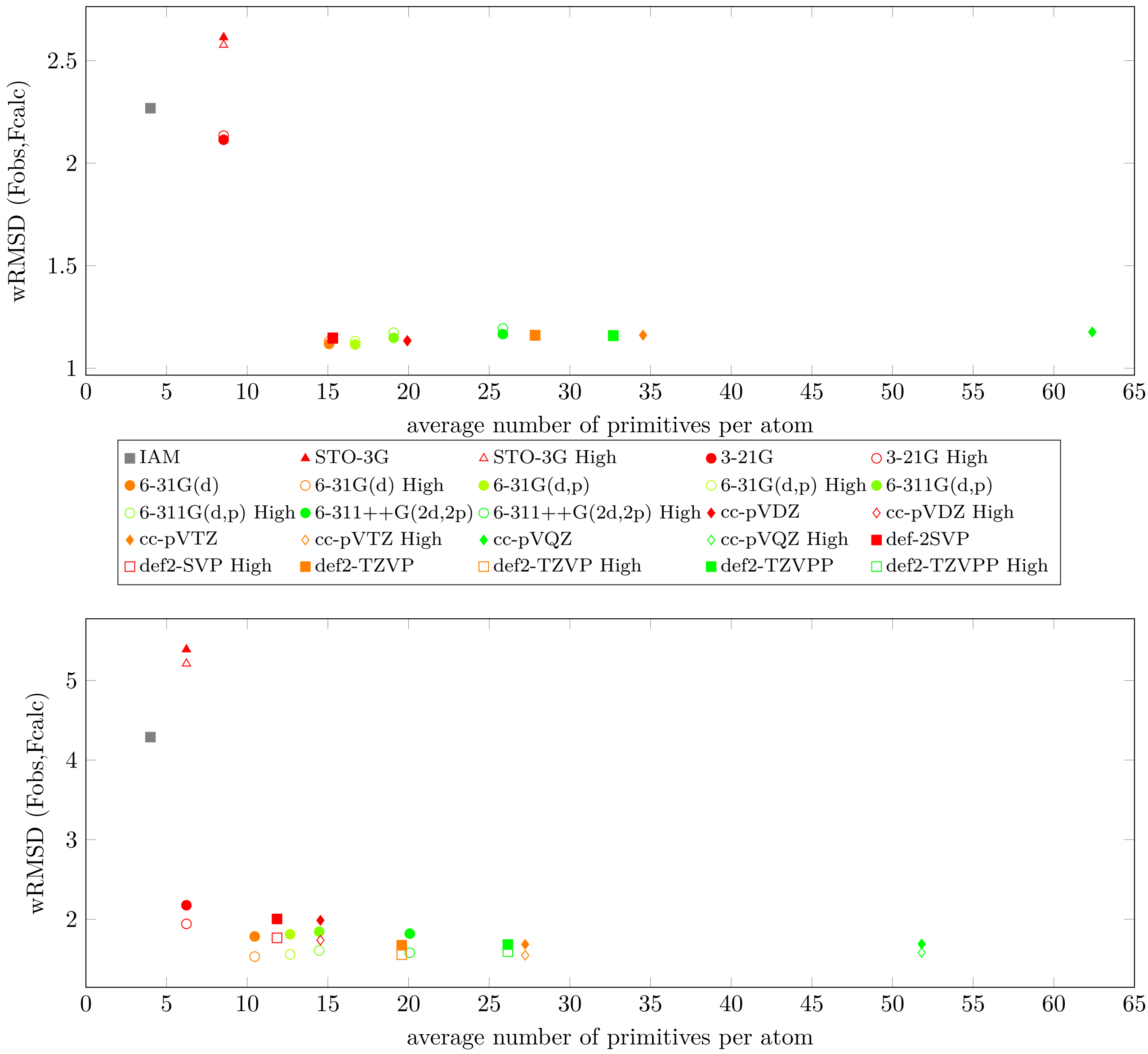


Figure 3.6: Plot of the wRMSD between $F^2{}_c$ and $F^2{}_o$ for L-alanine (top) and ammonium borate (bottom) against the average number of primitive basis functions used for the wavefunction calculation. Solid data points denote normal integration accuracy, outlined data points the corresponding high accuracy refinements. Basis sets of the same family have the same shape of data points, except for the IAM, which was arranged at 4 primitives per atom, as the modern scattering factors are calculated based on 4 Slater functions that were pre-tabulated.

wavefunction calculation and partitioning software. The final results of all comparisons are shown in Figure 3.6. The switch from STO-3G to 3-21G in the region of lowest number of basis functions, which has the same number of basis function in these test cases, significantly improves the refinement for both molecules. In the case of $(NH_4)_2B_6H_6$, the wRMSD is reduced by approx. 60% when using 3-21G compared to the STO-3G calculation. Also 3-21G performs much better than the IAM, which was not the case for STO-3G.

Interestingly, in the case of L-alanine already the 6-31G and def2-SVP calculations with less than 20 primitives (s-type orbital = 1 primitive, p type orbital = 3 primitives, d type orbital = 6 primitives, f type orbital = 10 primitives, g type orbitals = 15 primitives) per atom already produce wRMSDs comparable to higher basis sets. The higher basis sets show some fluctuation around these values between 1.1 and 1.2, but no obvious trend. The only observable trend is that higher grids do not improve the result, but to the contrary in some cases even increase the disagreement.

Only in the case of $(NH_4)_2B_6H_6$ the higher integration grids improved the refinement. Here the improvement was quite significant, but decreases with the number of basis functions. This behavior might result from the high symmetry of the structure, which reduces the number of unique reflections which can be used during the refinement. Here it seems the use of higher basis sets and grids is beneficial, however the cc-pVQZ refinement is not better than def2-TZVP, cc-pVTZ or 6-31G(d). Only STO-3G, 3-21G, def2-SVP and cc-pVDZ perform noticeably worse in terms of the wRMSD at the higher grid choice.

It can be summarized that not the highest level of theory, but rather a mixture of medium sized basis set together with accurate, but not exaggerated, integration grids yields in the best performance in terms of computational cost and accuracy balance. The effect these choices have on the quality of the refinement besides accurate description of the measured intensities has to be evaluated separately. Also it should be noted that these tests were only performed on relatively light elements and usign only on DFT functional. Therefore a more extensive systematic study needs to be conducted following these first insights.

3.1.3 *Publication:* Accurate Crystal Structures and Chemical Properties from NoSpherA2

NoSpherA2, the interface between quantum chemical software and *Olex2* was written by me, with significant feedback by Horst Puschmann to facilitate the automatic generation and use of *.tsc* files, presented in section 3.1.2.1. The first implementation using *Tonto* for the wavefunction calculation and Hirshfeld partitioning was performed based on some routines from Dylan Jayatilaka that could write ASCII archives of the scattering factors of the individual atoms, that were then merged together into a *.tsc* file and read by *Olex2*. The refinement engine itself, based on the *cctbx*, [50, 435] was implemented by Oleg Dolomanov. The *cctbx* is a collection of crystallographic algorithms, tools and open source software that can be used for various purposes, ranging from numerical conversion between different units, calculation of molecular structural parameters like distances and angles based on an input model or the refinement of crystal structures and evaluation of the model accuracy and fit to the used experimental data.

Since later the need to implement interfaces to *ORCA*, *Gaussian* and *pySCF* arose, I also wrote independent code to convert the resulting files into the *Tonto*-readable format *.fchk*, which is native for *Gaussian09* based on code from Alessandro Genoni to transform the orbital coefficients into the correct order and normalization and calculate the density matrix from the *.wfn* file. All remaining pieces of the file, including geometry, basis set information, atom weights, basis set shell structure was freely written by me. However, *Tonto* proved to be incorrect for heavier elements, especially atom types heavier than Kr. Therefore, the partitioning and transformation of electron density of a wavefunction was rewritten, inspired by *Tonto*, but completely standalone, using the openly available *numgrid* which is a C++ wrapper of the spherical Lebedev rules. [341, 429] This program was derived using the available code for handling wavefunctions and deriving properties from them as done in *cuQCrT*, especially the density. This software will be introduced in chapter 3.3. The *Thakkar* basis [427, 428] was implemented up to Rn and optimizations in terms of speed and interfaces and compatibility done and file handling programmed in *Olex2*. Later the plotting of derived properties like Laplacian, ESP, ELI etc. was implemented based on *cuQCrT* code (see also section 3.3.2). The source code for the *NoSpherA2* executable is available at *https://github.com/FlorianKleemiss/wfn2fchk*.

One interesting new finding of this publication is the disagreement of the IAM when describing CaF_2. HAR can more accurately describe the measured intensity of low order reflections of this salt. When looking at residual and deformation densities, it becomes obvious that the non-spherical description is a significant improvement. As discussed in the publication, the question already arose in the literature whether this was the case, but to this date no clear conclusion was drawn and the textbook knowledge still completely neglects any contribution different than packing of integer charged spherical particles.

All Figures, Tables and Schemes of this publication were made by me. The first draft of the complete text was written by me, which was iterated between Horst Puschmann, Simon Grabowsky and me several times to yield the final version shown here. All refinements, data evaluation and CIFs were prepared by me. Measurement of the CaF_2 dataset was performed by me on the in-house diffractometer.

Reprinted with permission from Royal Society of Chemistry 2020, DOI. 10.1039/D0SC05526C.

Chemical Science

ROYAL SOCIETY OF CHEMISTRY

EDGE ARTICLE

Cite this: DOI: 10.1039/d0sc05526c

All publication charges for this article have been paid for by the Royal Society of Chemistry

Accurate crystal structures and chemical properties from NoSpherA2†

Florian Kleemiss,[a] Oleg V. Dolomanov,[b] Michael Bodensteiner,[c] Norbert Peyerimhoff,[d] Laura Midgley,[d] Luc J. Bourhis,[e] Alessandro Genoni,[f] Lorraine A. Malaspina,[a] Dylan Jayatilaka,[g] John L. Spencer,[h] Fraser White,[i] Bernhard Grundkötter-Stock,[j] Simon Steinhauer,[j] Dieter Lentz,[j] Horst Puschmann *[b] and Simon Grabowsky *[a]

Received 6th October 2020
Accepted 6th November 2020

DOI: 10.1039/d0sc05526c

rsc.li/chemical-science

The relationship between the structure and the properties of a drug or material is a key concept of chemistry. Knowledge of the three-dimensional structure is considered to be of such importance that almost every report of a new chemical compound is accompanied by an X-ray crystal structure – at least since the 1970s when diffraction equipment became widely available. Crystallographic software of that time was restricted to very limited computing power, and therefore drastic simplifications had to be made. It is these simplifications that make the determination of the correct structure, especially when it comes to hydrogen atoms, virtually impossible. We have devised a robust and fast system where modern chemical structure models replace the old assumptions, leading to correct structures from the model refinement against standard in-house diffraction data using no more than widely available software and desktop computing power. We call this system *NoSpherA2* (Non-Spherical Atoms in Olex2). We explain the theoretical background of this technique and demonstrate the far-reaching effects that the improved structure quality that is now routinely available can have on the interpretation of chemical problems exemplified by five selected examples.

1. Introduction

1.1 The importance of crystallography for science

Single-crystal X-ray structure determination is arguably the most important analytical technique available to chemists since it alone can reveal the three-dimensional structure of matter cheaply, routinely and – above all – unambiguously.

The impact of this technique on the scientific developments in chemistry, biology, materials sciences, engineering, and physics cannot be overstated. To date, 26 Nobel Prizes in medicine, chemistry, and physics have been awarded to more than 50 researchers directly associated with crystallography.[1] Some of the milestones of science are based on single-crystal X-ray diffraction, including the elucidation of the structures of DNA[2] and graphene.[3] Large investments are made to push diffraction techniques for three-dimensional structure determination to new limits. These include the construction of X-ray synchrotron facilities and X-ray free-electron lasers[4] as well as the exploration of possibilities to employ other radiation types like from neutron spallation sources[5] and electron diffractometers.[6] Unfortunately, crystallographic methods and software development have not kept up with hardware development. The vast majority of structure refinements are still based on techniques that make use of one crucial simplification that was introduced in the early days of crystallographic refinement: the

---

[a]*Universität Bern, Departement für Chemie und Biochemie, Freiestrasse 3, 3012 Bern, Switzerland. E-mail: simon.grabowsky@dcb.unibe.ch*

[b]*OlexSys Ltd, Durham University, South Road, Durham, DH1 3LE, UK. E-mail: horst@olexsys.org*

[c]*Universität Regensburg, Fakultät für Chemie und Pharmazie, Universitätsstr. 31, 93053 Regensburg, Germany*

[d]*Durham University, Department of Mathematical Sciences, South Road, Durham, DH1 3LE, UK*

[e]*Bruker France, 4 Allée Lorentz, Champs-sur-Marne, 77447 Marne-la-Vallée cedex 2, France*

[f]*Université de Lorraine & CNRS, Laboratoire de Physique et Chimie Théoriques (LPCT), UMR CNRS 7019, 1 Boulevard Arago, F-57078 Metz, France*

[g]*University of Western Australia, School of Molecular Sciences, 35 Stirling Highway, WA 6009, Perth, Australia*

[h]*Victoria University of Wellington, School of Chemical and Physical Sciences, Wellington 6012, New Zealand*

[i]*Rigaku Europe SE, Hugenottenallee 167, 63263 Neu-Isenburg, Germany*

[j]*Freie Universität Berlin, Institut für Chemie und Biochemie, Anorganische Chemie, Fabeckstr. 34/36, 14195 Berlin, Germany*

† Electronic supplementary information (ESI) available: CIFs of all refinements, further electron-density analysis of compound HQ-CO_2, theoretical background of non-spherical refinements in NoSpherA2 and the tsc file format, graphs and tables documenting the validation procedure described in Section 3. CCDC 2034385–2034389, 2035147 and 2035148. For ESI and crystallographic data in CIF or other electronic format see DOI: 10.1039/d0sc05526c

Independent Atom Model (IAM), in which atoms are assumed to be separate, non-interacting spherical entities.[7]

Here, we show how an accessible generalized procedure in which quantum chemical calculations coupled with modern crystallographic software can solve a variety of chemical problems using single-crystal diffraction data of any kind.

1.2 How it used to work

X-rays interact with the electrons in a crystalline material, and this interaction gives rise to measurable diffraction. While most of the electrons are centered on the atoms themselves, some electrons are involved in chemical bonding and are therefore not located where they would be in non-interacting, spherical atoms, which are assumed in any standard refinement. This means that the diffraction pattern obtained from the experiment is due to the real distribution of the electrons in the compound of interest, not due to spherical distributions.

The diffraction pattern consists of thousands of unintelligible spots on hundreds of images. To make sense of it, it is necessary to resort to a model of the molecule that will give rise to the measured diffraction pattern. This model building is central to any technique based on diffraction, and it is this final model that we call a 'crystal structure'.

Calculating the expected diffraction pattern from this model requires a certain amount of computing power – and this is where the approximation comes in. Calculations are a lot easier if a spherical distribution of electron density around each atom is assumed. This IAM approximation works very well since most of the diffraction is due to the electrons on the atoms themselves – and the quality of diffraction data that could be obtained in the early days of the field was itself rather limited.

Today, X-ray crystallography finds itself in the strange position where the real diffraction pattern arising from the interaction of X-rays with electrons in their real positions can be measured with high precision and accuracy, and yet we still use an approximation in our models that leads to assuming the wrong positions of these electrons.

1.3 What we gain when we leave the century-old spherical approximation behind

In every chemical compound, electrons are involved in bonding of one kind or another – and in traditional X-ray crystallography, this has been ignored. Only if we model this non-sphericity we can gain deep insights into intra- and intermolecular chemical bonding. Detailed and accurate information on chemical bonding from the diffraction experiment is of major importance, for example, for materials design,[8] catalysis,[9] drug design,[10] and chemical textbook education.[11] Non-spherical atomic electron density features are highly significant for hydrogen atoms, which have only a single valence electron, and therefore the relative error caused by a spherical core approximation is most severe. Using X-rays, hydrogen atom positions can be obtained with the same accuracy and precision as afforded by neutron diffraction experiments, but only if the non-spherical nature of the electron distribution can be accounted for.[12] Improved structural information is important in many areas, including the understanding of thermodynamic properties such as heat capacities,[13] interaction or lattice energies,[14,15] or the development of force fields.[16]

1.4 Non-spherical structure refinement

There are two strategies by which we can introduce atomic non-sphericity: either we actually *refine* the electron density (experimental electron density determination) or we find *theoretical* ways of introducing the shape of the electron cloud that is associated with each atom and then use that shape when we refine the structure against the diffraction pattern.‡ This is what this work is about; and to avoid any possible confusion with electron density refinements, we will use the term 'non-spherical structure refinement' when referring to using non-spherical atomic form factors.

In a standard structure refinement, where atoms are treated as independent of each other (IAM), the atomic form factors have been calculated theoretically from the spherical electron-density distribution of isolated atoms in the past and are now available in tabulated form.[25] Hence, refinements are fast and convenient, but they neglect chemical bonding, as discussed above. Non-spherical structure refinements are either based on approximate pre-calculated and averaged multipole populations or theoretical wavefunctions, which are tailor-made for the compound under investigation. In both cases, non-spherical bonded-atom electron densities are produced and are then Fourier-transformed to produce non-spherical atomic form factors. However, only in the latter case, the theoretical chemical-bonding information is readily available after the refinement and can be chemically interpreted.

1.4.1 Multipole-based databank approaches. Multipole parameters can be calculated theoretically from synthetic structure factors of model compounds and stored in tables according to the atom type defined in its chemical environment (Invarioms[26] or UBDB[27]). Alternatively, such multipole databanks can be constructed from averaged experimental electron densities (ELMAM[28]). Multipole parameters are then transferred from the databank to the compound under investigation, and non-spherical atomic form factors are calculated for use in the refinement on the fly from the transferred multipole populations. Although such refinements produce, *e.g.*, better structural parameters for hydrogen atoms,[29] they are currently restricted to organic and bio-organic compounds, for which the number of atom types is manageable. A simplification to only dipole level for chemical bonds is now also implemented in the *ShelXL* software.[30]

1.4.2 Hirshfeld atom refinement. The non-spherical structure refinement method called Hirshfeld atom refinement (HAR)[31,32] is central to this study. Starting from the atomic positions obtained from a standard X-ray structure, a molecular wavefunction is calculated using quantum mechanical calculations and then dissected into atomic electron density functions (the Hirshfeld atoms, see Fig. 1) using Hirshfeld's stockholder partitioning scheme.[33] The resulting Hirshfeld atoms are never stored in tables, but a Fourier-transformation

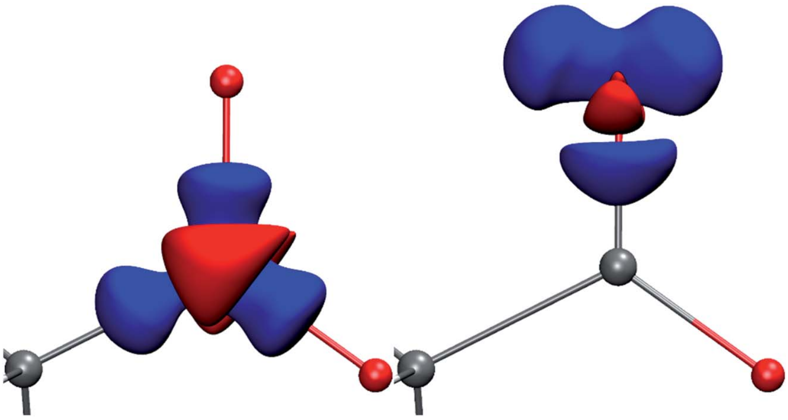


Fig. 1 Deformation Hirshfeld densities[12] for the carbon (left) and oxygen (right) atoms in the carboxylate group of Gly-L-Ala, *i.e.* difference between the spherical atomic electron density used in the IAM and the non-spherical Hirshfeld atom density used in HAR (IAM *minus* HAR). Red = negative, blue = positive. Isovalue = 0.17 eÅ^{-3}.

of their electron density is carried out on the fly to produce the related atomic form factors used in the least-squares refinement. This will produce an improved structure, and the procedure is then repeated – a new wavefunction and new atomic form factors are calculated after each refinement cycle – until convergence is reached. This makes HAR the most accurate of the non-spherical structure refinement methods.[34]

1.4.3 Current implementations of HAR and their limitations

The first implementation of HAR was based on the software *Tonto*,[35] and at least two early interfaces to the procedure have been built: *lamaGOET*[36] and *HARt* in conjunction with *Olex2*.[37] Both tools are limited by the capabilities of *Tonto*. Hirshfeld atoms in *Tonto* are not optimized for elements heavier than Kr, which can introduce large numerical errors for heavy elements. They also rely on *Tonto* as the actual crystallographic refinement engine, which it was not primarily designed to be. Vital features such as restraints, treatment of special positions, partial occupancies, twinning, solvent masking, and reliable CIF output are missing.

HAR requires the repeated calculation of a molecular wavefunction, which restricts its applications even further. The overall process can be slow because of the repeated quantum-mechanical step. One approach is to combine HAR with libraries of extremely localized molecular orbitals (ELMOs),[38] which has resulted in the HAR-ELMO method.[39] While this is very fast, it relies on the availability of pre-calculated molecular orbitals and is therefore not suitable for general use, but is highly relevant for the refinement of proteins which consist of a fixed subset of 20 amino acid building blocks.

Periodic network compounds could not be handled previously because molecular wavefunctions are used – a problem discussed in ref. 40 by analyzing periodic wavefunctions in combination with stockholder partitioning.

Disordered compounds cannot be handled by *Tonto*, and there are no apparent plans that this feature will be implemented in *Tonto* in the foreseeable future.

1.5 NoSpherA2 enables new possibilities for HAR

In this work, we present a new implementation of HAR in *Olex2*: *NoSpherA2* (Non-Spherical Atoms in Olex2). It decouples HAR from *Tonto* and instead makes use of the *Olex2* Graphical User Interface (GUI) as well as of the fully-featured *olex2.refine* least-squares refinement engine. All modeling options (including restraints, constraints, disorder modeling, solvent masking, *etc.*) thus become accessible to HAR. Besides, the *NoSpherA2* development also focussed on the accurate description of core electrons and spin states for heavy elements. This opens HAR to all those users already familiar with standard structure determination procedures and extends its use to include almost all classes of compounds, so that they can be routinely determined.

We present here three different compound classes that could not previously have been refined with HAR – but exhibit important chemical-bonding questions:

- disordered structures (both occupational and conformational disorder).
- Structures in highly symmetric space groups with special positions.
- Structures with heavy elements next to very light elements.

We have refined representative structures of each class using *NoSpherA2* as summarized in Table 1.

1.5.1 Disordered structures. A search in the Cambridge Structural Database[41] shows that 27% of all crystal structures are affected by disorder. Hence, it is of utmost importance to be able to extend HAR toward the treatment of disordered compounds. Here, we distinguish between occupational and conformational disorder.

Occupational disorder. Occupational disorder relates to a part of the crystalline compound not being present in every unit cell. This happens regularly in host–guest systems,[42] *e.g.* in loaded metal–organic[43] or covalent-organic frameworks.[44] In the crystalline sponge method,[45] the host framework is well known, but it is the structure of the mostly disordered guest which is to be elucidated. Understanding host–guest interactions in such systems widely used as storage or analysis tools for smaller

Table 1 Summary of problems solved by the NoSpherA2 implementation of HAR and the exemplary structures shown in this work. Additionally, possible fields of application that benefit from this are given

Problem	Example	Field of Application
Occupational disorder	$(C_6H_6O_2)(CO_2)_{0.854}$	Supramolecular chemistry, host–guest systems, MOFs/COFs,
Conformational disorder	$C_{10}H_{10}N_4F_2$	Protein crystallography, solvent disorder, macromolecular crystallography
High symmetry network compounds	$CaF_2/(NH_4)_2B_6H_6$	Network compounds, crystal design, inorganic structure refinement
Compounds containing heavy metals	$OsH_6(PC_{12}H_{19})_2$	Inorganic and metalorganic compounds, catalytic complex understanding, metal–metal interactions

molecules relies on an accurate description of the location and geometry of both the host and the, very likely disordered, guest molecule. We have chosen the CO_2-hydroquinone clathrate as an example of how the guest-molecule position and partial occupation can be refined accurately in HAR.[46,47]

Conformational disorder. In biological macromolecular crystallography, disorder is omnipresent in the main molecules as well as in the solvation sphere. This kind of disorder consists of different conformations being spatially overlaid. For their description, it does not matter if they are caused by dynamic effects or are static. However, dynamic disorder is the most frequent in protein crystallography and could imply many different conformations that need to be modeled. Therefore, an extension of HAR to macromolecular chemistry crucially depends on the ability to treat the complex disorder in different disorder groups.[39] Here, we demonstrate how non-spherical atomic form factors from different molecular wavefunctions corresponding to different molecular conformations can be combined for an accurate HAR of a tetrahydropyrido[2,3-*b*]pyrazine derivative, a compound class of interest for drug development.[48]

1.5.2 Structures in highly symmetric space groups with special positions. Inorganic materials often consist of solid ionic network compounds. Until now, this entire compound class has been excluded from HAR. We have chosen a textbook ionic salt (fluorite, fluorspar, CaF_2) and the molecular salt ammonium hexahydrohexaborate $(NH_4)_2B_6H_6$ to demonstrate the ability of *NoSpherA2* in *Olex2* to refine such compound classes using HAR for the first time. We show that, even for such simple compounds, an analysis of improved geometrical parameters plus a theoretical wavefunction perturbed by the respective crystal field yields deep and perhaps even surprising insights into bonding phenomena.

The textbook notion that species such as CaF_2 consist of spherical ions has rarely been questioned, although there is evidence that either a multipole or a neutral-atom model may describe such compounds better than the conventional ionic model.[49,50] Electron and γ-ray diffraction have shown that there is non-sphericity of the valence electron density in KCl, LiF, and MgO.[51,52] Specifically for Ca^{2+}, the formally empty d-orbitals can be partially populated to add substantial covalent bonding character to the cation–anion interactions, as, for example, in the putative Zintl-phase of CaSi[53] or the calcocenes.[54] Here, utilizing HAR, we show that there is significant bond-directed non-spherical valence density present in the crystal structure of CaF_2.[49]

The bonding in polyhedral boranes and borates cannot be explained by 2-center-2-electron Lewis models, but various other concepts such as three-dimensional aromaticity must be invoked.[55] Therefore, there are many diffraction- and electron-density-based studies on borates.[56] Here, we have chosen $(NH_4)_2B_6H_6$ because of its fundamental character and symmetric crystal packing, but also because in previous studies the quantum-crystallographic description involving the hydrogen atoms was ambiguous.[57] However, if the potential of closo-hexaborates for hydrogen-storage applications[58] is to be fully exploited, the hydrogen atom geometric and bonding parameters must be determined accurately and precisely, which will be demonstrated in this paper.

1.5.3 Structures with heavy elements. Another area where the accurate determination of hydrogen atom parameters is of tremendous importance is transition-metal catalyzed hydrogen activation and hydrogenation.[59] Specifically, the structures of heavy atom hydrides are of interest, but it is extremely challenging to determine parameters of hydrogen atoms bonded to heavy elements accurately by X-ray crystallography.[12] Already by itself, the description of the heavy element in heavy-element containing species is challenging enough from both the crystallographic and the quantum-chemical point of view.[39,60,61]

Here, we report how the methodological progress manifested in *NoSpherA2* allows the successful non-spherical refinement of the osmium atom and the accurate determination of hydrogen atom parameters in bis(diisopropylphenylphosphine) hexahydridoosmium, $OsH_6(PC_{12}H_{19})_2$,[62] referenced against results from neutron diffraction of the same compound.[63] Limits of the X-ray diffraction experiment and its resolution truncation are also revealed and discussed.

2 NoSpherA2

NoSpherA2 brings wavefunction calculations, non-spherical atom partitioning, subsequent atomic form factor calculation, and finally least-squares refinement together under the umbrella of the freely available *Olex2* software.[64] It is presented here utilizing the HAR non-spherical structure refinement method, but it is compatible with other flavors of (X-ray and electron) non-spherical structure refinement as well.

Besides the *NoSpherA2* method developments discussed below, the advantages gained within HAR by switching from the *Tonto* to the *olex2.refine* refinement engine are manifold and significant. It allows HAR to access the use of restraints and constraints, hydrogen atom riding models, correct crystallographic description of special positions, refinement of partial occupancies, twinning models, solvent disorder treatment *via* BYPASS, different choices of weighting schemes, and many more options. In *olex2.refine*, refinements are carried out in F^2, not in F as previously done in *Tonto*. Advantages of the use of the *Olex2* GUI include the automatic generation and validation of crystallographic information files (CIFs), the generation of maps and plots, and the straightforward generation of completed molecules or clusters of symmetry-related molecules on the screen, which are then used in the refinement as input.

2.1 Fundamental concepts

The diffraction pattern and the electron density associated with the geometry of the measured crystal are related *via* a mathematical method, the Fourier transform. More precisely, the Fourier transform of the electron-density distribution ρ_{xyz} in the unit cell is the structure factor F_{hkl} depending on the Miller indices *hkl* (eqn (1)). The F_{hkl} are complex numbers, and the square of their amplitudes are proportional to the measured Bragg-reflection intensities identified by Miller indices *hkl* representing vectors $\vec{h}$ in the reciprocal lattice. The structure factor is obtained as a finite sum involving the atomic form factors (also known as 'atomic scattering factors') $f_{\mathrm{hkl},j}$ of each atom j in the unit cell. The atomic form factors are, in turn, derived from the atomic electron density distributions *via* a Fourier transform. N_{atoms} is the number of atoms in the unit cell, $\vec{r}_j$ is the position vector of atom j, and T_j is the Debye–Waller factor of atom j, that is, an exponential function involving the atomic displacement parameters of atom j. These values appear in the equation describing the crystallographic model (eqn (1)):

$$F_{hkl} = \sum_{j}^{N_{\mathrm{atoms}}} f_{hkl,j} e^{2\pi i \vec{h} \vec{r} j} \sum_{j} T_j\left(\vec{h}\right) \tag{1}$$

In the classical Independent Atom Model (IAM), the atoms are considered independent non-interacting entities, and their electron densities are spherical functions depending only on the atomic type. The resulting form factors are listed in tables.[25] In a model that takes interatomic interactions into account, the atomic electron densities – translated to the origin – *are no longer spherical functions* and *are dependent on the geometry of the whole molecule*. Therefore, the corresponding atomic form factors are no longer real, but complex-valued functions of the Miller indices.

The idea behind the flexible approach to non-spherical structure refinement that underlies *NoSpherA2* is to provide these complex-valued form factors of each atom within the unit cell *via* a table given in a file provided from plugin software (*.tsc*-file).[65] The underlying total electron density can be provided by quantum-mechanical calculations or other sources.

For molecules that are related by symmetry within the unit cell, only the form factors of atoms within the asymmetric unit are required, since the geometric transformation from an atom in this asymmetric unit to an equivalent atom in the unit cell corresponds to a transformation of the Miller indices of the corresponding non-spherical atomic form factor. Therefore, it is necessary to pay attention that all Miller indices resulting from transformations for all equivalent atoms are provided in the *.tsc*-file.

Moreover, the least-squares minimization used in non-spherical structure refinement requires not only information about the non-spherical form factors themselves but also about their partial derivatives. To avoid the time-consuming process of additional *.tsc*-files at close-by configurations, for which the numerical densities of the individual atoms vary only very little under tiny changes of their atom positions, we neglect these tiny variations in our calculations of the partial derivatives. The validity of this approximation will be discussed in more detail in a forthcoming publication.[66] The mathematical details of this general and flexible approach as well as the precise format of the *.tsc*-files are discussed in the ESI.†

2.2 NoSpherA2 GUI and interface

Fig. 2 summarizes the interplay of different steps of the non-spherical structure refinement and related software programs interfaced by *NoSpherA2*. At first, the input for the refinement is selected in the *Olex2* graphical user interface (GUI) (Fig. 2) from a grown structure if necessary. The atomic coordinates of this structure are then transferred to the chosen quantum-mechanical software in form of a *.cif* or *.xyz* file. At present, *ORCA*,[67] *Gaussian*,[68] and *Tonto*[35] are interfaced. The software *ORCA*, which is free of charge for academic use, has been tested most extensively here and has been used for all the examples in this work. The wavefunction output (*.wfn* or *.wfx* formats) is transferred to the program of choice that performs the Hirshfeld atom partitioning and the subsequent atomic form factor

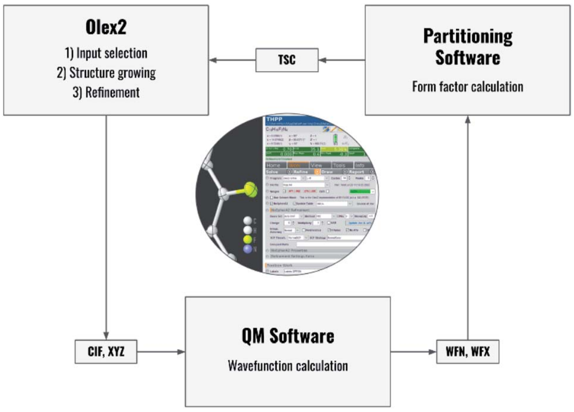


Fig. 2 Flowchart of the process behind HAR controlled by *NoSpherA2* and a visual representation of the GUI for *NoSpherA2* inside *Olex2*.

calculation. Supported programs for this are *Tonto* or *NoSpherA2* itself, whereby *Tonto* is only used for validation purposes in this study, and all the examples are based on *NoSpherA2*. Finally, the atomic form factors are handed to *olex2.refine* in *.tsc* format (see the previous subsection and ESI†) for regular least-squares refinement in *Olex2*. The entire cycle visualized in Fig. 2 can be iterated manually or automatically until convergence in all refined parameters within less than 1% of their standard uncertainties is reached. This automatically iterated refinement is consistent with the standard definition of Hirshfeld Atom Refinement (HAR).[32]

Every aspect of the entire process can be controlled directly from within the *Olex2* GUI. Depending on the quantum mechanical package used, different choices affecting the molecular wavefunction calculation will be available. The required basis sets are provided by *Olex2* and contain all relevant elements of the periodic table. A detailed description of the available options is provided in Section 5.

2.3 Dealing with disorder

For compounds including occupational disorder, the only methodological difference to previous HARs in *Tonto* is that *olex2.refine* allows the refinement of partial occupancy parameters. For conformational disorder, *NoSpherA2* recognizes disorder parts previously defined inside the *Olex2* GUI. Subsequently, individual molecular wavefunctions are calculated separately for every conformer and individual _part*.tsc* files are written for every wavefunction, which will in the end be combined to a single _*total.tsc* file for refinement. If there are more than two disorder parts, they can be grouped so that in each group the parts add up to 100% occupancy (in a semicolon-separated list in the *Olex2* interface, *e.g.* "1–5"). Molecular wavefunctions are then calculated for every combination of parts between the disorder groups. This tool is essential for protein crystallography, which we will test in a separate forthcoming study.

2.4 Open-shell wavefunctions: multiplicity

The handling of open-shell wavefunctions was introduced in *NoSpherA2*, so that non-spherical atomic form factor calculations become possible for any spin state as found, for example, in high- or low-spin transition metal complexes. As long as the unrestricted or even multi-configurational wavefunction, *e.g.* after a CASSCF calculation, is presented in *.wfn* or *.wfx* format, *NoSpherA2* will read the information and calculate the electron density based on all fully or partially occupied molecular orbitals and produce the *.tsc* file accordingly.

2.5 Dealing with heavy elements

To address the heavier elements of the periodic table, the inclusion of all-electron basis sets covering such atoms was necessary (x2c basis sets of Pollak and Weigend[69]). Also, the DKH2-relativistic 2-component Hamiltonian[70] approach was made accessible in combination with any HF or DFT method selected (the DKH2 method should only be used with the x2c basis sets).

The calculation of atomic form factors is different from previous HAR implementations, and this has the biggest impact on heavy elements. Integrable grids are calculated with an adaption of *numgrid*[71] which uses a Lindh–Malmqvist–Gagliardi radial description[72] and a modern implementation of the spherical Lebedev quadrature procedure.[73,74] Becke partitioning[75] with a hardness factor of 3 is applied; and for the Hirshfeld partitioning[33] spherical atomic electron densities calculated from Thakkar densities,[76] represented by Slater-type functions, are used. More details are provided in Section 5.

2.6 Visualization of derived properties and functions

Olex2 natively includes engines for the generation of two- and three-dimensional maps and plots of functions for the representation of residual electron densities. In the course of the implementation of the *NoSpherA2* software, we have significantly extended the plotting options from dynamic and static deformation electron density and Laplacian of electron density maps to properties that can only be obtained with wavefunction information. This includes the electron localizability indicator ELI,[77] electrostatic potentials, molecular orbitals, and the non-covalent interaction NCI index.[78] Therefore, all pictures in this paper except for Fig. 1 and 4(b) are generated with the *Olex2* software.

3 Validation of HAR in NoSpherA2

The multi-temperature X-ray and neutron-diffraction data sets of L-Ala and Gly-L-Ala represent a well-established benchmark set of structures already used in previous tests and validations of HAR.[32,39,79] Here, five X-ray datasets of Gly-L-Ala and three X-ray datasets of L-Ala were refined at HF/6-311G(d,p) using six different refinement techniques each (Table 2).

In all models and datasets, the hydrogen atom positions and displacement parameters were freely refined. In all HAR models, hydrogen atoms were refined anisotropically. Fig. 3 shows an indicative comparison between the residual density distributions after a standard (IAM) refinement and a HAR with model vi. After IAM, distinctive positive residual densities remain on the covalent bonds and in the lone pairs of the oxygen atoms. After HAR, the bonding and lone-pair densities have been fully accounted for by the non-spherical atomic form factors (compare deformation Hirshfeld density representations in Fig. 1).

Table 2 Models used during validation with combinations of selected software and parameters to show the validity of all steps independently

ID	Type	Program	QM	Partitioning	Weighting scheme
i	IAM	olex2.refine	—	—	$1/\sigma^2(F^2)$
ii	HAR	Tonto	Tonto	Tonto	$1/\sigma(F)$
iii	HAR	NoSpherA2	Tonto	Tonto	$1/\sigma^2(F^2)$
iv	HAR	NoSpherA2	ORCA	Tonto	$1/\sigma^2(F^2)$
v	HAR	NoSpherA2	ORCA	Tonto	Shelxl-type
vi	HAR	NoSpherA2	ORCA	NoSpherA2	$1/\sigma^2(F^2)$

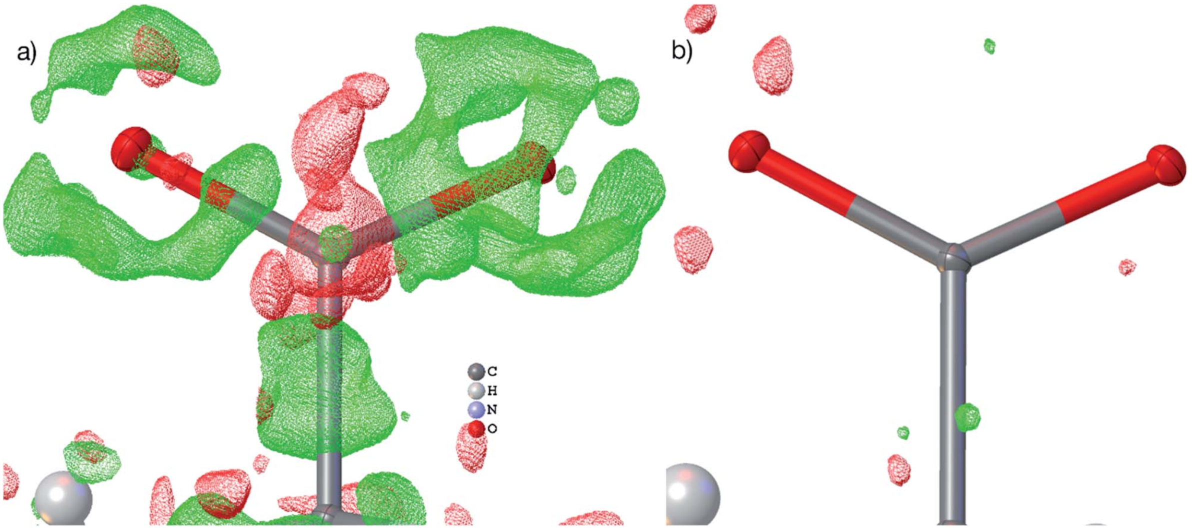


Fig. 3 Visualization of the residual density distribution of the carboxylate group in L-Ala after IAM (a) and HAR, model vi (b). The residual density was calculated with *Olex2* from fcf files and plotted on a grid of 0.05 Å with an iso-value of 0.10 $eÅ^{-3}$ (green = positive, red = negative). In the IAM plot, residual density regions of a different functional group that obstructed the view onto the carboxylate were manually removed.

The tested models iii to vi introduce an increasing amount and combination of parameters that distinguish the new *NoSpherA2* HARs from the traditional *Tonto* HARs (model ii). Summarizing the validation results (Tables S2–33 and Fig. S5–10, ESI†), all HARs present the same accuracy and precision relative to the neutron-diffraction results regardless of the way the model was generated. The exchange of the least-squares refinement in *Tonto* with *olex2.refine* through the *.tsc* files provides statistically identical results, with all differences being far below the measurement uncertainty. However, the results cannot be numerically identical because of the differences in the procedure discussed in Section 2 (such as refinement in F^2 instead of F, different Becke grids used, *etc.*). We find that the iteratively updated *ShelXL*-type weighting scheme is advantageous concerning the derivation of ADPs when compared to neutron-diffraction results. The combination of *ORCA* and *NoSpherA2* also seems to produce a slightly closer agreement with the neutron-diffraction results for all parameters compared to the *Tonto*-derived results. Certainly, the results are produced much faster with *ORCA* and *NoSpherA2* (183 seconds) relative to *Tonto* (884 seconds, Gly-L-Ala at 23 K, 6 CPUs), and additionally, the grid density is higher in *NoSpherA2*. The final refinements used for producing the results shown in Section 4 are based on model vi as this is indicated as the most promising combination of settings: HAR in *NoSpherA2* using *ORCA*-wavefunctions and *NoSpherA2*-partitioning with a *Shelxl*-type weighting scheme.

4 Results and discussion

Table 3 provides an overview of the different nature of all five compounds discussed in Section 4 concerning symmetry, resolution and data/parameter ratios. It further describes the data quality as well as the extent and success of the refinements, focusing on a comparison of the IAM with the HAR results. There are two general observations:

Table 3 Refinement indicators using IAM and HAR. More details are given in the ESI, Table S1

Selected parameters	$(C_6H_6O_2)\,(CO_2)0.854$	$C_{10}H_{10}N_4F_2$	CaF_2	$(NH_4)_2(B_6H_6)$	$OsH_6(PC_{12}H_{19})_2$
Space group	$R\bar{3}$ (trigonal)	$P2_1/n$ (monoclinic)	$Fm\bar{3}m$ (cubic)	$Fm\bar{3}m$ (cubic)	$P2_1/n$ (monoclinic)
No. of unique reflections	857	2975	96	364	13 109
d_{min}/Å (radiation source)	0.58 (Mo)	0.70 (Mo)	0.40 (Ag)	0.40 (Mo)	0.58 (Mo)
No. of parameters (constraints/restraints), IAM *vs.* HAR	51(3/0) *vs.* 70(1/0)	204(7/0) *vs.* 258(1/4)	3(0/0) *vs.* 3(0/0)	11(0/0) *vs.* 11(0/0)	421(0/0) *vs.* 636(0/6)
Level of theory	PBE/def2-TZVPP DKH2-PBE/x2c-TZVPP				
Charge/multiplicity	0/1	0/1	+18/1	+6/1	0/1
R_{int}/%	1.73	5.36	6.73	2.41	5.82
R1 (IAM)/%	3.26	3.89	1.31	1.84	2.11
R1 (HAR)/%	1.45	2.14	1.14	0.95	1.92
$\Delta\rho$ (IAM)/$eÅ^{-3}$	0.581/−0.179	0.402/−0.313	0.690/−0.476	0.179/−0.216	1.128/−1.093
$\Delta\rho$ (HAR)/$eÅ^{-3}$	0.227/−0.254	0.175/−0.216	0.686/−0.415	0.119/−0.066	1.167/−0.938

- *R*-factors and residual densities are significantly lower for HAR in comparison to IAM refinements.
- *R*-factors after HAR are very low, regardless of R_{int} values, space group, resolution of measured data or the presence of disorder.

4.1 Disordered structures

4.1.1. Occupational disorder in $(C_6H_6O_2)(CO_2)_{0.854}$ (HQ-CO_2). The central motif in the hydroquinone crystal structure is a void encapsulated by hydrogen-bonded rings of hexagonal topology. These voids are normally filled with guest molecules forming clathrate structures; in fact, it is difficult to keep hydroquinone guest-free.[80] Various guest molecules can be trapped and then be transported through the host structure, which leads to various applications of hydroquinone clathrates.[46,47,81,82]

Many studies focus on carbon dioxide inside the hydroquinone voids for fuel and energy science, but the occupancy (or filling ratio) of the voids is unclear.[46,47,81,83] In the example of the HQ-CO_2 compound shown here, the occupancy of the entity in the void was refined in HAR to be precisely 0.854(2) (Fig. 4(a)). The experimental details are provided in Section 5.

Not only was it possible to determine the occupancy of CO_2 precisely, but all hydrogen-atom positions and anisotropic displacement parameters were obtained accurately and precisely. There is one symmetry-independent hydrogen bond that closes (by symmetry) the two six-membered rings that encapsulate the void (Fig. 4). HAR-refined geometric details of this hydrogen bond identify it as short and strong: $d(O\cdots O) = 2.6805(3)$ Å, $d(O–H) = 0.963(6)$ Å, $d(H\cdots O) = 1.724(6)$, $a(O–H\cdots O) = 171.7(6)°$. Three void channels are intersecting the unit cell (see Fig. S1†). Having accurate and precise hydrogen-atom positions from HAR, the volume of each void could be estimated with the software CrystalExplorer[84] to be 68.5 Å^3.[85] In total, 16% of the unit cell are guest-accessible void volumes. The Hirshfeld surface[86] of the CO_2 molecule encloses a volume of 44.4 Å^3, which means that the CO_2 molecule fits into this particular void but is presumably not closely bound. Further analysis of the Hirshfeld surface using the property d_{norm} mapped onto it (Fig. 4(b)) reveals that there are no contacts closer than the sum of the van-der-Waals radii of the atoms in CO_2 and the atoms of the host structure; in fact, d_{norm} is positive throughout the entire range. This implies that there are only weak van-der-Waals and dispersion forces between host and guest, enabling the guest molecule to travel through the void channels. Nevertheless, the CO_2 molecule is not dynamically disordered in this crystal structure – it is just not always present in the void.

4.1.2 Conformational disorder in $C_{10}H_{10}N_4F_2$ (THPP). The THPP crystal structure provides an example of a conformational disorder where two different disorder parts are present within the same disorder group (compare Section 2.3). In HAR, the disorder could not only be resolved unambiguously, but the hydrogen atoms in the major disorder component (88.2(5)%) could even be refined anisotropically (Fig. 5(a)). The THPP molecule was split into two disorder parts, *i.e.* different

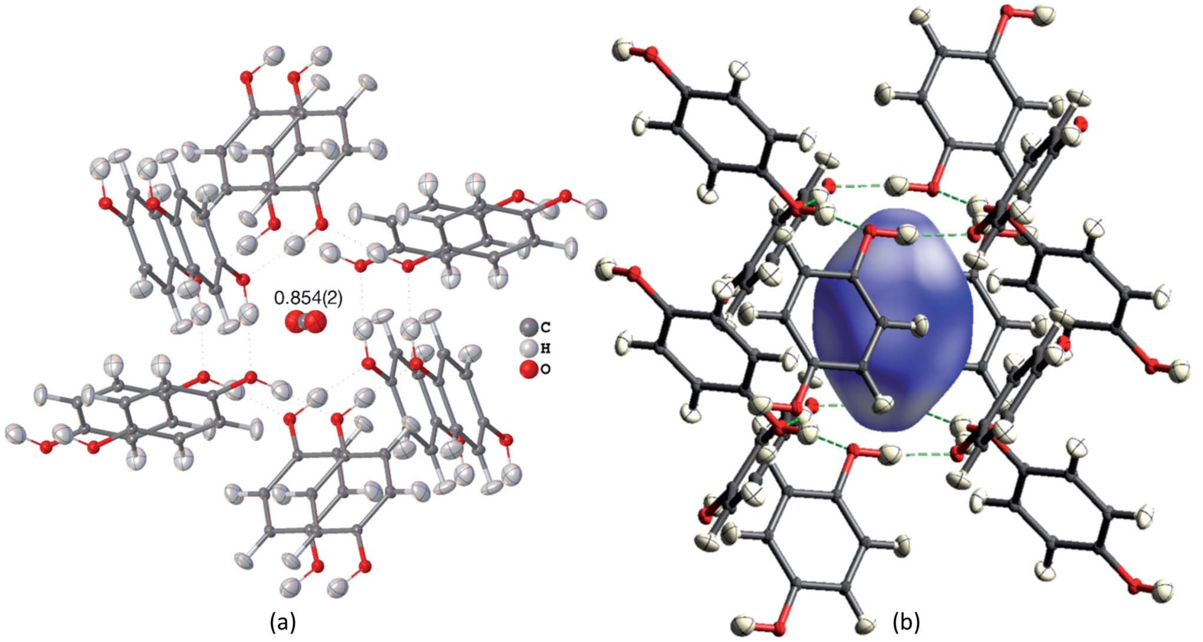


Fig. 4 HAR-refined hydroquinone-CO_2 clathrate structure (HQ-CO_2) with anisotropic displacement parameters depicted at 80% probability level. The cluster of the guest CO_2 molecule with the 12 surrounding HQ molecules encapsulating CO_2 inside a void is shown as used in the wavefunction calculation underlying the non-spherical form factor generation. (a) Final refined geometry and partial occupation number of CO_2. (b) Hirshfeld surface representation of CO_2 inside the void mapped with the property d_{norm}. Color scale from 0.08 (white) to 0.77 (blue). Generated with CrystalExplorer.[87]

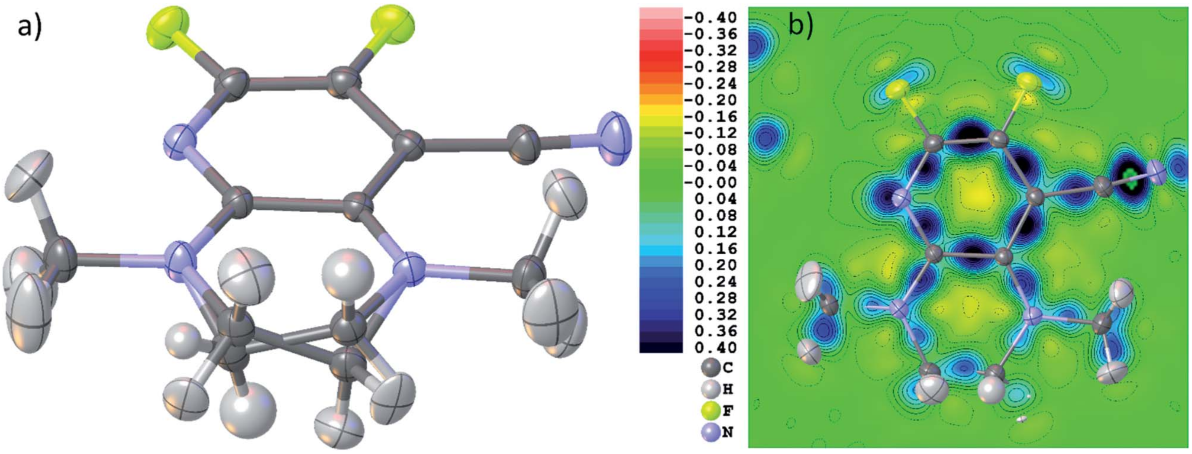


Fig. 5 (a) Final HAR geometry after disorder treatment of THPP. (b) Dynamic deformation density distribution in the main molecular plane of the molecule. Color scale legend in eÅ^{-3}. Atomic anisotropic and isotropic displacement parameters at 80% probability level.

conformers, including the two nitrogen atoms next to the methylene CH_2 groups to allow for the calculation of different atomic form factors of these nitrogen atoms in different chemical environments while fixing their positions to be the same. Some angle and ADP restraints had to be used on the methylene groups of the minor disorder component (11.8(5)%) and the hydrogen atoms of this minor part were refined isotropically with bond length constraints to the major part.

Fig. 5(b) shows a dynamic deformation density map in the main molecular plane, *i.e.* the difference electron density of the HAR and the IAM including the effect of refined atomic displacement parameters. The map confirms that all the details of chemical bonding can be analyzed from this HAR disorder refinement, also in the disordered region where several displacement ellipsoids overlap. Moreover, the lone pair regions at nitrogen and fluorine atoms are accurately shaped.

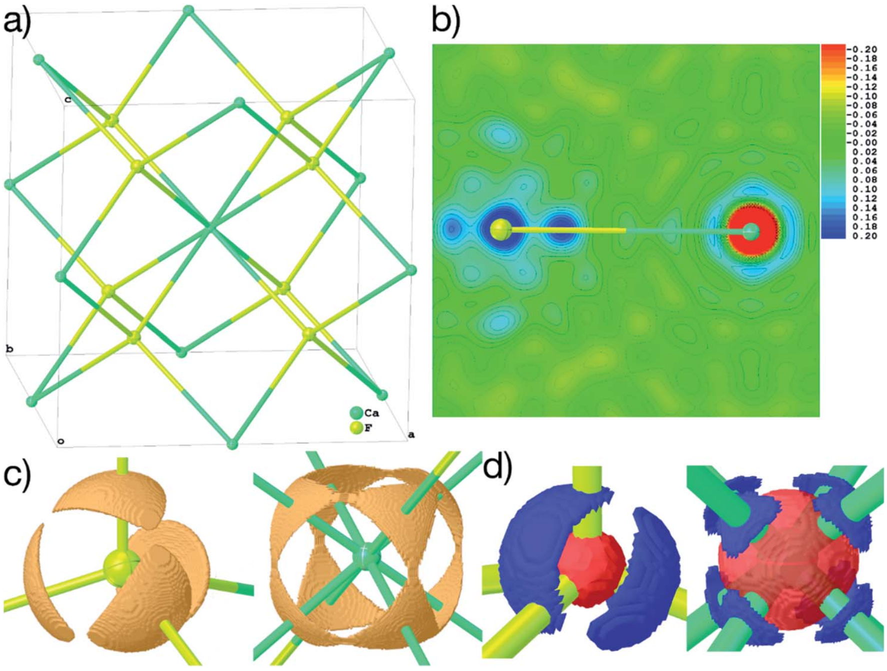


Fig. 6 (a) Structure of the explicit cluster used for the calculation of the wavefunction of CaF_2 during HAR, coinciding with a unit cell. (b) 2D-plane of dynamic deformation density, color scale in steps of 0.02 eÅ^{-3}. (c) 3D-isosurfaces of the ELI-D with iso-values of 1.835 (F) and 1.910 (Ca). (d) 3D-isosurfaces of the atomic deformation Hirshfeld densities at isovalues of ±0.411 (F) and ±0.088 (Ca) eÅ^{-3}. All maps are calculated on a 0.01 Å grid and plotted using *Olex2*. Displacement parameters at 80% probability level.

4.2 Structures in highly symmetric space groups with special positions

4.2.1 Fluorite CaF_2. Since periodic-boundary conditions of solid-state quantum mechanical programs are not yet available in *NoSpherA2*, we tested many explicit clusters of Ca^{2+} and F^- ions for the wavefunction calculations in *ORCA*, up to several hundred ions large to ensure a proper description of the network polarizing the asymmetric unit. In comparison to the large clusters, it turned out that the minimal cluster, which consists of the completed coordination sphere (octahedron) of a central calcium ion and of the completed coordination spheres (tetrahedra) of the eight adjacent fluoride ions (Fig. 6(a)), is sufficient to accurately determine the displacement parameters and the properties of the ions obeying the symmetry of the system. The cluster consists of 21 ions, a total charge of +18, and multiplicity 1. Both ions Ca^{2+} and F^- are located on special positions in such a way that all their coordinates are fixed (distance = 2.3603 Å) and only two displacement parameters and the scale factor are refinable (see Table 3). This also means that in this case, the wavefunction for HAR does not need to be updated during the refinement since the atomic positions do not change.

Fig. 6(b)–(d) show unambiguously that the description of CaF_2 as an ionic salt with spherical ions is incorrect. There are significant charge concentrations in the deformation density maps (Fig. 6(b) and (d)) and charge localization in the electron localizability (ELI-D) map (Fig. 6(c)) directed along the Ca–F interactions. Hence, non-spherical, bond-directed valence electron density distribution caused by polarization, charge transfer, and electron-density deformation play a significant role in the bonding in CaF_2. In an analysis of the wavefunction within the framework of QTAIM[87] the atomic charges imply significant charge transfer from F^- to Ca^{2+} (+1.74 e (Ca) and −0.87 e (F)), there is a Ca–F bond critical point with an electron-density value of 0.22 eÅ^{-3} and a Laplacian value of 4.43 eÅ^{-5}, as well as a delocalization index[88] of 0.14, which can be interpreted as the partial bond order.[49] It remains to be clarified by more detailed follow-up studies to which extent covalency, d-orbital population, or core deformation interplay to support the ionic framework of the crystal structure. However, we note that covalency and ionicity are not necessarily opposing forces but might be two sides of the same coin.[89,90]

The maps in Fig. 6(b)–(d) are based on the theoretical cluster wavefunction, and represent the input for the non-spherical structure refinement, most directly represented by the deformation Hirshfeld density plots in Fig. 6(d), the difference between the IAM and HAR densities used in the refinement. This information can be used according to ref. 91 to directly show the significance of the non-spherical signal in the X-ray diffraction data, supporting similar findings by electron and γ-ray diffraction.[51,52] For this purpose, in Fig. 7 the difference between the calculated F_c and the measured F_o structure factors in both models IAM and HAR is plotted relative to their resolution; weighted by their standard uncertainty σ in Fig. 7(a) and

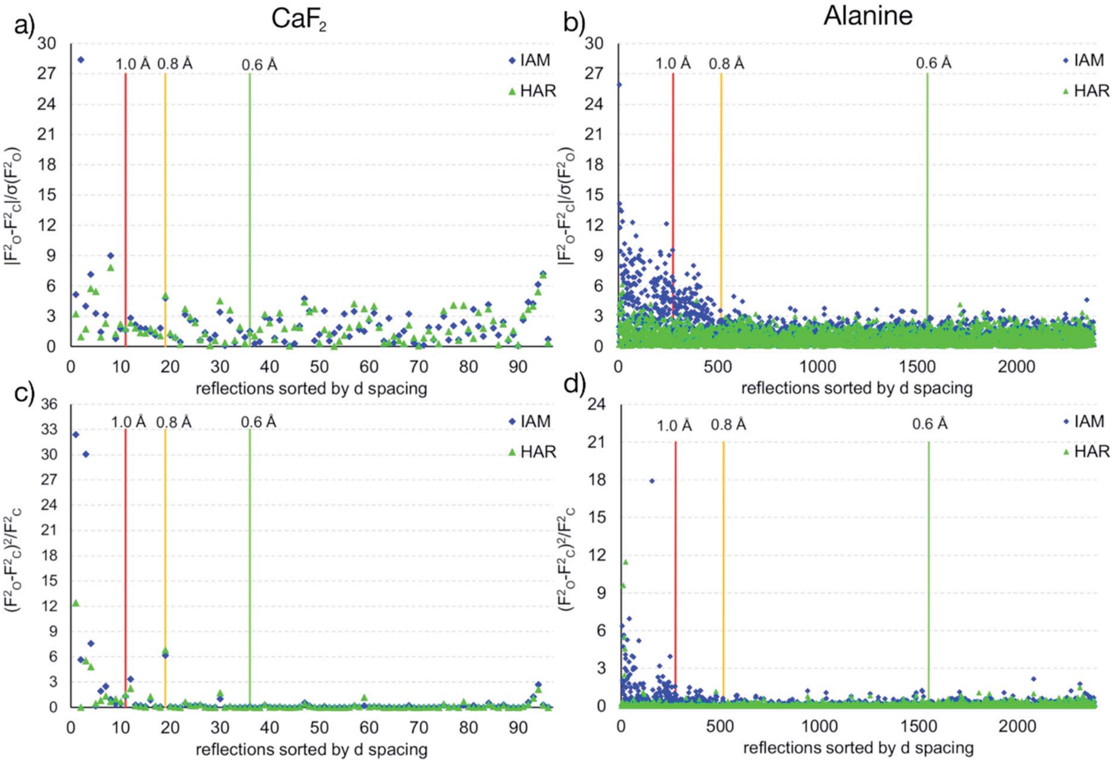


Fig. 7 Differences of observed and calculated structure factors *versus* resolution for CaF_2 and L-Ala for the IAM and the HAR models.

(b), or by their absolute magnitude in Fig. 7(c) and (d). The non-spherical valence density signal is expected to be more relevant for the low-order reflections. Consequently, for a compound such as alanine with many strongly covalent bonds, the IAM shows large discrepancies between the measured and modeled structure factors below *ca.* 0.8 Å resolution, whereas HAR does not show such model insufficiencies probed by the experiment (Fig. 7(b) and (d)). For CaF_2, the same systematic effect is not as strongly pronounced, but it is clearly present for the 8 to 10 lowest-order reflections out of a total of 96 reflections in this data set. This shows that HAR can reveal the degree of non-sphericity by model comparison with the X-ray diffraction experiment directly, which was so far believed to be only possible for the more precise convergent-beam electron diffraction experiment. In turn, the model of spherical ions is not suited to describe the ions in CaF_2.[49,50]

The shortcomings of the IAM model impact on the refined parameters because the neglected non-sphericity must be absorbed by the weighting scheme and the atomic displacement parameters. The coefficients of the weighting scheme are $a = 0.0217$ and $b = 0.3133$ in the IAM, while the coefficients in HAR are reduced to $a = 0.0175$ and $b = 0.0607$. This trend of significantly smaller weighting scheme factors is observed for all other refinements, as well. The two refinable U_{iso} values are 0.00337(5)/0.00495(9) Å^2 for Ca/F in the IAM, which change to 0.00325(7)/0.00488(9) Å^2 if the simple $1/\sigma^2$ weighting scheme is used. The differences in HAR are slightly smaller: 0.00334(4)/0.00505(6) Å^2 *vs.* 0.00328(5)/0.00502(7) Å^2.

4.2.2 Ammonium hexahydrohexaborate $(NH_4)_2B_6H_6$. For $(NH_4)_2B_6H_6$, many different symmetric and asymmetric, large and small clusters were tested for the wavefunction calculation. As for CaF_2, a minimal cluster that obeys the crystallographic symmetry is sufficient for an accurate and precise HAR of the borate anion of interest.

Here, a six-fold positively charged cluster with the central borate octahedron neighbored by all 8 ammonium ions in the

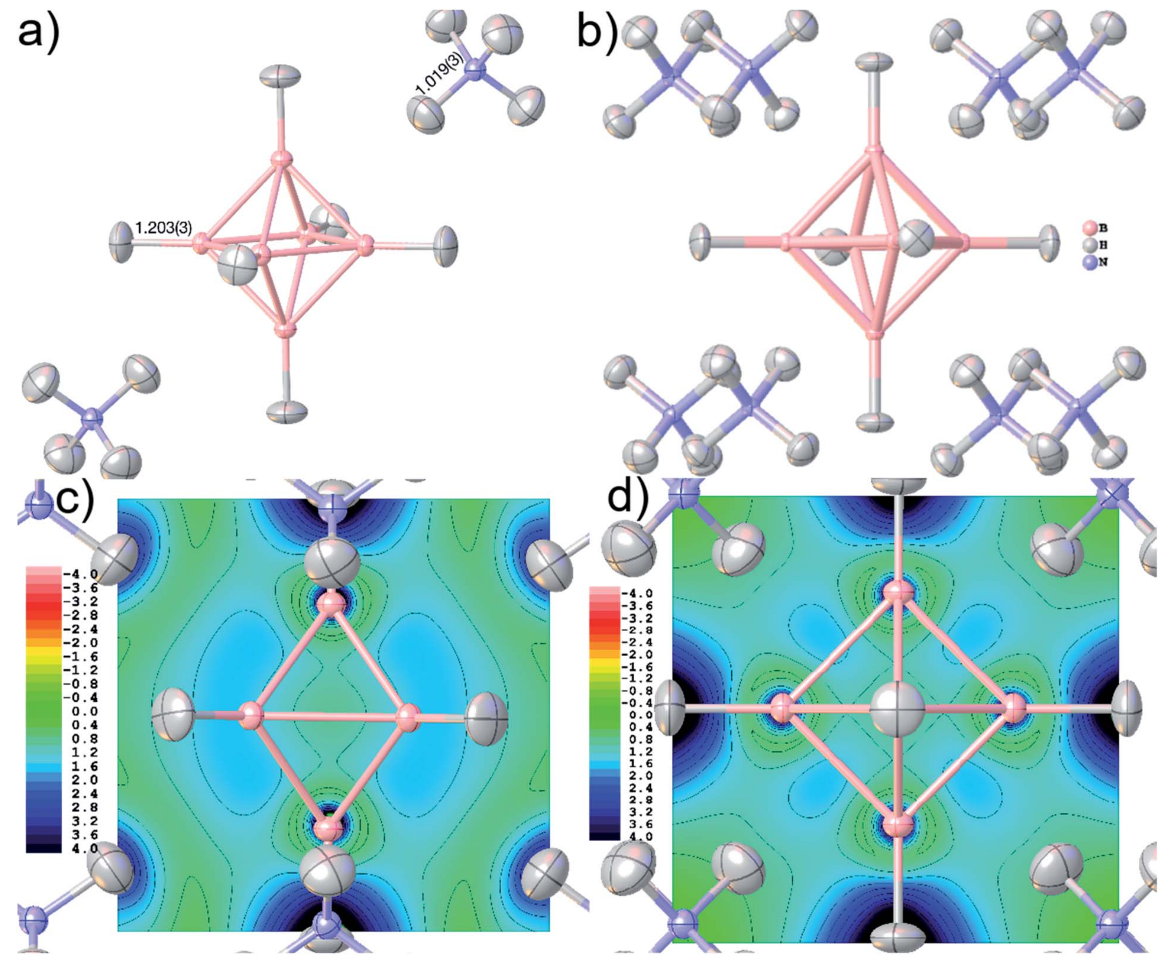


Fig. 8 Final geometry and anisotropic displacement parameters at 80% probability level of ammonium borate $(NH_4)_2B_6H_6$ after HAR, showing (a) the formula unit with the two symmetry-independent refined bond distances in Å, and (b) the molecular cluster used in the wavefunction calculation. Electron localizability indicator ELI-D at contour intervals of 0.4 in a cut-plane (c) intersecting two pairs of opposite BBB faces, and (d) in the central boron square.

first coordination sphere was used (Fig. 8(b)). Preliminary tests with single-zeta basis sets resulted in inferior residual densities and X–H distances (X = N,B); only after using a triple-zeta basis set the refinement improved considerably compared to IAM. The refinement using PBE/def2-TZVPP resulted in a significant drop in R-value and residual density (see Table 3) and gave accurate N/B–H distances of 1.019(3) Å and 1.203(3) Å, respectively (Fig. 8(a)). Reported values for N–H distances in ammonium ions from neutron diffraction have an average value of 1.021 Å and a standard deviation of 0.037 Å.[92] For any borane or borate clusters, we find only two single-crystal neutron diffraction studies in the literature yielding an average terminal B–H bond distance of 1.195 Å with a standard deviation of 0.009 Å over 14 symmetry-independent B–H bonds.[93,94] For comparison, the IAM fails to produce similar X–H distances: N–H 0.836(7)/B–H 1.078(6) Å.

Accurate X–H distances are crucial for the derivation of properties such as charge concentrations and localizations related to the special bonding situation in boranes and borates. As an example of possibilities inside *Olex2* for bonding analysis after HAR, Fig. 8(c) and (d) show two different ELI-D maps as cut-planes intersecting the B–B–B faces and in the central B–B–B–B square. The ELI-D maxima are clearly outside the bond axes and are delocalized around the boron polyhedron, which is in line with previous theoretical calculations that show the ELF/ELI polyhedron being dual/complementary to the structural $B_6H_6^{2-}$ polyhedron, within the theory of three-dimensional aromaticity.[57,95,96]

4.3 Structures with heavy elements

The refinement of hydrogen atom parameters in heavy metal hydrides is one of the most challenging aspects of X-ray crystallography. In fact, not only is the diffraction pattern dominated by the heavy element,[97] but truncation errors of the Fourier series of the structure factors that are limited by resolution also spatially occur in regions where the hydrogen atoms are located. Therefore, to overcome these drawbacks, it is necessary to collect both very high-quality low-order data to capture the hydrogen-atom signal and high-resolution data to reduce truncation errors.[98] However, the experimental X-ray diffraction data of compounds containing heavy elements are very often affected by systematic problems such as significant absorption and radiation-damage effects.[39,99] Here, we test to which extent a very sophisticated theoretical electron-density model underlying the refinement (see Section 2.5) can help to interpret the diffraction pattern of the osmium hexahydride $OsH_6(PC_{12}H_{19})_2$.

In all HARs reported in Table 4, the Os–H distances were refined freely, while some restraints on the hydrogen atom anisotropic displacement parameters were applied, and one hydrogen atom was refined isotropically. Extinction correction was also applied during the refinements. Calculations were performed using the B3LYP, M06-2X, and PBE DFT functionals, always with the DKH2 relativistic extension and the basis set x2c-TZVPP. This series was repeated for a subset of the reflections (pruned at $d = 0.7$ Å), as there is a significant rise in R_{int} for data beyond this resolution. In Table 4, the *R*-value, the residual density minimum and maximum values, and the timing for the QM step are compared. More details and model differences are deposited as CIFs with the paper.

The geometry and the displacement parameters of the refinement using the full resolution are shown in Fig. 9. The sum of covalent radii[100] of Os and H is 1.61 Å, which is significantly longer than the distances resulting from the HAR refinement (Fig. 7(b), average 1.554 Å, average standard uncertainty 0.014 Å). However, the corresponding neutron-diffraction experiment yielded longer Os–H bonds, on average 1.649 Å.[62] The isotropic IAM refinement, in turn, yielded much shorter Os–H bond lengths (1.510 Å). When the resolution cut-off from 0.58 to 0.7 Å was used, the average HAR-derived bond distance and the average standard uncertainty remain constant at 1.555 Å and 0.015 Å. This means that the significant difference in *R*-value and maximum residual-density value located at the Os core between the two resolutions (Table 4) are unrelated to the Fourier truncation error and do not impact on the hydrogen atom treatment. It is unclear whether the advantage of having more information from higher resolution data or the disadvantage of compromising on the overall data quality by including more high-resolution data prevail over the other.

In summary, the HAR results for those hydrogen atom parameters in $OsH_6(PC_{12}H_{19})_2$ that are bonded directly to the Os atom are improved relative to the IAM results but are still significantly less accurate and less precise than those for compounds involving only lighter elements. To understand whether the insufficient match with the neutron-diffraction derived Os–H bond lengths is caused by problems in the HAR methodology or complications of X-ray diffraction experiments on heavy-element containing species, we calculated and refined a theoretical structure factor set of $OsH_6(PC_{12}H_{19})_2$ based on the neutron-derived geometry at the same level of theory as used in HAR. A dynamic set of structure factors was obtained by a combination of *ORCA*, the *.tsc* routine in *NoSpherA2*, and the *.fcf* routine in *Olex2* up to the same resolution (0.58 Å) as the experimental structure factor file. In addition to the information on the displacement parameters, the theoretical structure

Table 4 Comparison of HARs of $OsH_6(PC_{12}H_{19})_2$ using different DFT functionals, extended with the DKH2 relativistic method and the basis set x2c-TZVPP

Funct. res.	B3LYP full	B3LYP 0.7 Å	M06-2X full	M06-2x 0.7 Å	PBE full	PBE 0.7 Å
R1/%	1.94	1.20	1.93	1.19	1.92	1.19
$\Delta\rho$/eÅ^{-3}	1.210/−0.708	0.592/−0.317	1.210/−0.681	0.581/−0.318	1.167/−0.686	0.583/−0.317
t of QM step/s	639		1335		579	

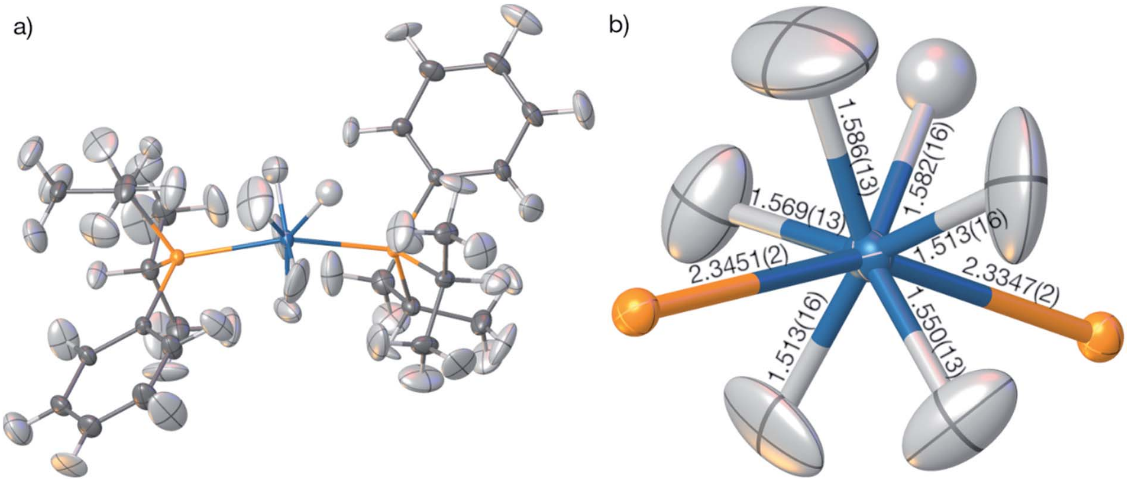


Fig. 9 (a) HAR-refined molecular structure of $OsH_6(PC_{12}H_{19})_2$ with the DKH2-PBE/x2c-TZVPP model at full resolution. (b) Coordination geometry of the Os atom with the freely refined distances to the six nearest hydrogen and two nearest phosphorus atoms. All displacement parameters are displayed at 80% probability level.

factors include information on anomalous dispersion based on the Sasaki table.[102] Uncertainties were set to 0.001 of the calculated intensities.

The structure was solved from scratch on the basis of the theoretical structure factors. A subsequent IAM refinement resulted in shortened Os–H distances (av. 1.574 Å) in comparison to the input structure (av. 1.649 Å). This means that the IAM model is insufficient and cannot produce the input parameters even for theoretical data. This is reflected in the residual density distributions depicted in Fig. 10(a) where unmodelled Os–H bonding density and overestimated Os core density is visible at levels as high as 0.58 and −0.26 eÅ^{-3}. After HAR at DKH2-PBE/x2c-TZVPP with high integration accuracy, coordinates and atomic displacement parameters agree exactly to the last digit with the input values (see CIFs deposited as ESI†), and the residual electron density has vanished (0.003 eÅ^{-3}, see Fig. 10(b)). These results imply that the experimental X-ray diffraction data and not the HAR model are the reason for the inaccurate determination of the Os–H bond lengths discussed above.

It is worth noting that the $OsH_6(PC_{12}H_{19})_2$ measurement used here is not of especially inferior quality, but rather represents a standard measurement as it is nowadays routinely obtained for service measurements of coordination and organometallic compounds. Therefore, we use this example to have a closer look at the problem of truncation effects caused by limited resolution because it was shown recently in ref. 98 that even at resolutions as high as $d = 0.20$ Å, core and outer-core electron-density distributions of a mercury hydride cannot be reproduced at all from structure factors, regardless of the sophistication of the quantum-crystallographic model. Fig. 10(c) shows a detailed view of the electron-density map obtained by the Fourier transform of the structure factor set calculated for the Os hydride at $d = 0.58$ Å resolution. The area around the nucleus is highly positive (in fact so highly positive that the values cause some trouble for the color mapping). Further out, two shells of highly negative electron density values (red and yellow) are separated by another highly positive shell (blue). These are the so-called Fourier truncation ripples, and it is worth remembering how severe the effect is, with highly significant physically meaningless negative electron-density areas near the core of the heavy element. The third shell of positive electron-density values (light blue) is still as high as 5 to 10 eÅ^{-3}. It is located at a distance of about 1.5 Å from the Os nucleus and therefore overlaps with the hydrogen atom electron densities, which makes the localization of hydrogen atoms bonded to heavy elements and the refinement of their parameters so difficult.

These truncation ripples cannot be avoided due to the nature of the X-ray diffraction experiment and its resolution limitation. They can only be mitigated by extremely high resolution which is not available at home sources so far; and at synchrotron sources, radiation damage becomes much more likely to occur. Whenever there are small errors in the Fourier series of the calculated structure factors, the steep gradients between the highly positive and negative electron density regions cause huge errors and consequently large residual electron density effects. Therefore, a much better understanding of the physical background of effects such as absorption, fluorescence, radiation damage, anharmonic motion, thermal diffuse scattering, and other systematic effects impacting on heavy elements is needed, leading to more sophisticated correction procedures[39,97,103] and improved diffractometer hardware and software technology.

5 Settings/experimental part

5.1 Software details

To make the refinements convenient, a *NoSpherA2* Graphical User Interface (GUI, Fig. 2) was added to the "Refine" tab inside the *Olex2* GUI which appears once the *NoSpherA2* tick-box is

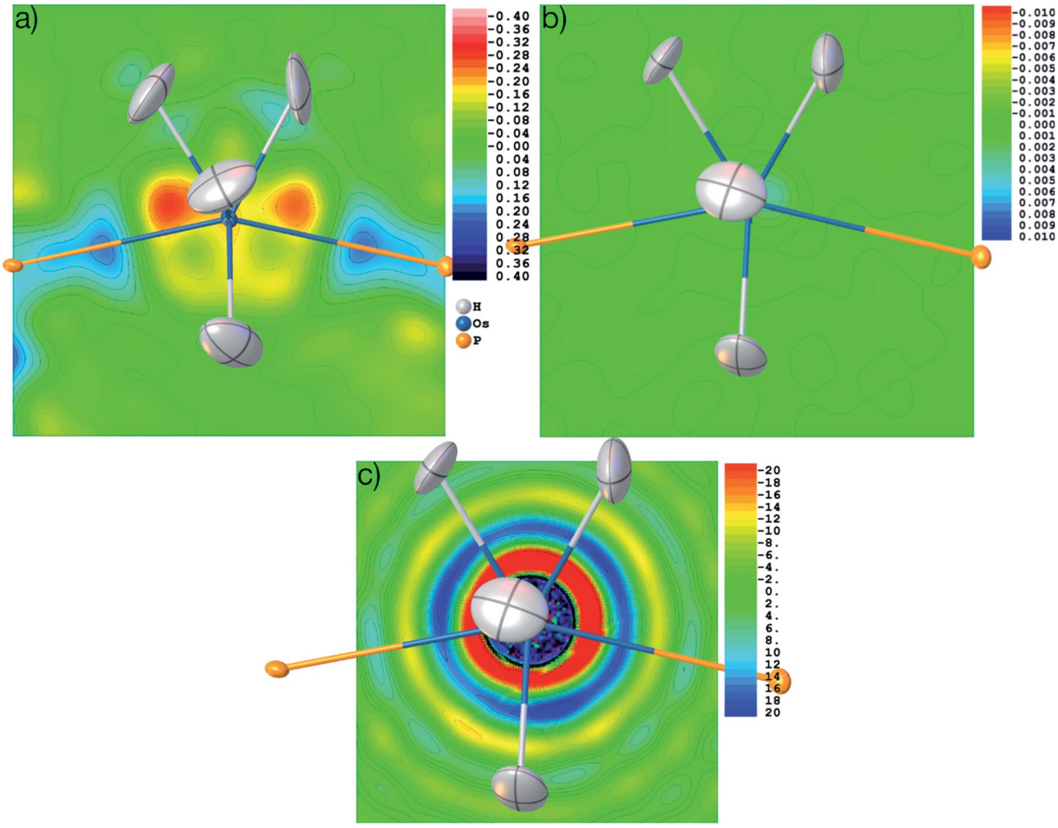


Fig. 10 Residual electron density after refinement against theoretical structure factors in (a) IAM model and (b) HAR model. (c) Electron density as the Fourier transform of the theoretical structure factors in the vicinity of the Os atom, showing Fourier truncation ripples. Three different contour intervals (see text). Color scale in eÅ^{-3}. Displacement parameters at 80% probability level.

activated. "Update Table" asks for the source of the *.tsc* file. For HAR, this means that either the QM software for the *.tsc* generation or an existing *.tsc* file must be chosen to be used in the least-squares refinement. The remaining user-specified options concern settings for all of the three major job steps. Basis sets and methods are pre-selected according to the QM software choice. PBE/def2-SVP is the minimal level for quick and yet reliable results. Most results in this paper are based on PBE/def2-TZVPP, because the PBE-GGA density seems to be well suited for the HARs,[101] whereas higher basis sets seem to be necessary for accurate atomic displacement parameters.[37] Hartree-Fock or hybrid methods such as B3LYP are also available. In principle, all kinds of levels of theory are possible in the input section of the QM code if the resulting electron density can still be evaluated on a Becke grid. The accuracy of Becke grids, both in the QM step and in the Hirshfeld atom partitioning, can be set. The application of a relativistic Hamiltonian is possible when activating the "REL" tick-box. More details on the treatment of heavy elements are discussed in Section 2.5.

The final refinement result is independent of the QM software used, but different QM programs provide different sets of features. If *Tonto* is chosen, a cluster of self-consistent Hirshfeld point charges and dipoles can be used to simulate the crystal field, and the radius of this cluster can be specified in the GUI. If *ORCA* is chosen, the resolution of identity (RI) and chain of spheres (COSX) approximations in *meta*-GGA functionals speed up the calculations without loss of accuracy of the refinement results. They are set by default, but the convergence threshold and convergence strategy can be controlled. Computational resources (number of CPUs and memory) can be allocated and might result in different gains of calculation speed in the different QM codes. In the future, we envisage the possibility to send calculations from the *NoSpherA2* GUI to a supercomputer infrastructure for further speed gains.

Crystallographic options concerning the hydrogen atom treatment in the refinement are only convenient switches since the least-squares procedure is controlled *via* the *Olex2* GUI in the same way an IAM *Shelxl* or *olex2.refine* refinement is controlled, including all the options such as riding models or restraints. Here, the tick boxes "H Aniso" will set all the hydrogen atoms to anisotropic displacement parameters at the start of the refinement, and "No Afix" removes all the previously fixed hydrogen atom parameters to ensure that HAR refines all parameters freely. "DISP" automatically adds $\Delta f'$ and $\Delta f''$ values for the anomalous dispersion correction from the Sasaki table[102] according to the wavelength specified in the input *.ins* or *.cif* files.

The calculation of grids is based on preselected levels of accuracy (low, normal, high, max; tick-box "Integr. Accuracy", see ESI† for more details), of which "normal" is usually sufficient, but "high" was used to reproduce the atomic electron density of the isolated osmium atom ($Z = 76$) with an integrated accuracy of better than 0.0001 e. According to the choice of integration accuracy used here for the atomic form factor calculation, a corresponding accuracy will also be selected for the wavefunction calculation in the QM software.

All HARs presented in the Results and Discussion part (Section 4) were carried out with the following settings: source of *.tsc*: *ORCA*; basis set: def2-TZVPP (except Os where x2c-TZVPP was used); method: PBE; relativistics: not used (except for Os); SCF Conv. Thresh.: NormalSCF; SCF Conv. Strategy: NormalConv; H Aniso: activated (if not mentioned otherwise); No Afix: activated; updated *Shelxl*-type weighting scheme: activated; automatic HAR; Integr. Accuracy: high; EXTI: not activated; DISP: activated. Cluster charges were not used for the simulation of the crystal field. Only for validation purposes and for the Os-containing compound, settings were varied. Other refinement details are shown in Table 3 as well as in the Crystallographic information files deposited with the Cambridge Structural and Inorganic Crystal Structure Database (CCDC-2034385 to 2034389, 2035147 to 2035148).

The setup for benchmarking the heavy element refinements was a 4-core 2.81 GHz hyperthreaded laptop with 16 GB RAM and no solid-state drives using 7 threads for ORCA. A significant part of the calculation was the evaluation and saving of relativistic integrals, which is similar for all functionals (approx. 125 s in these calculations). Apart from that, the PBE calculation was significantly faster than the MO6-2X calculation (161% longer), and still faster than the B3LYP calculation (13% longer). Furthermore, the results of the refinement using PBE are very similar to those of the more sophisticated functionals, so the use of PBE for all-purpose refinements is suggested, even when heavy elements are involved. In contrast, the application of relativistic corrections is imperative, and all-electron x2c basis sets lower than triple-zeta are not recommended. Therefore, all results discussed in Section 4.3 refer to DKH2-PBE/x2c-TZVPP HARs.

5.2 Origin of datasets and synthesis of compounds

For the validation part (Section 3), previously published X-ray structure factors of L-alanine (L-Ala)[104] and glycyl-L-alanine (Gly-L-Ala)[32] were used. For comparison, the results of neutron-diffraction studies at the respective same temperatures were used, from ref. 39 for L-Ala, and from ref. 105 for Gly-L-Ala. X-ray structure factors of the CO_2-hydroquinone clathrate (HQ-CO_2) were taken as deposited in the Cambridge Structural Database belonging to ref. 46, whereas those of the tetrahydropyrido[2,3-*b*]pyrazine derivative (THPP) stem from ref. 48.

The CO_2 molecule in HQ-CO_2 is located in a special crystallographic position. For the wavefunction calculations, we constructed a cluster of 12 hydroquinone molecules around the trapped CO_2 molecule enclosing the void (Fig. 4), but only a fraction of that cluster was refined as the crystallographic asymmetric unit, which consists of half a molecule of hydroquinone and half a CO_2 molecule ($Z = 9$, $Z' = 0.5$ formula units.

A high-resolution X-ray diffraction experiment of CaF_2, obtained as a small single-crystal from a fluorite mineralogical sample, was performed in-house on a Rigaku Synergy-S diffractometer equipped with a Hypix6000 detector at 100 K using Ag-K_α radiation. A single-crystal of ammonium hexahydrohexaborate $(NH_4)_2B_6H_6$ was synthesized for this study according to the procedure described in ref. 106 to yield the hexaborate anion as a sodium salt and subsequently yielding $(NH_4)_2B_6H_6$ after aqueous workup at pH 10 with ammonium chloride in solution. Single crystals were obtained by evaporation of the solvent after filtration. It was measured to high resolution using a Rigaku Synergy diffractometer equipped with a Pilatus 300 K detector at 100 K using Mo-K_α radiation. The compound bis(diisopropylphenylphosphine) hexahydridoosmium, $OsH_6(PC_{12}H_{19})_2$, was synthesized according to ref. 84 a suitable single crystal was measured to medium resolution on a Rigaku SuperNova EosS2 diffractometer with a CCD detector at 120 K using Mo-K_α radiation. Further crystallographic and measurement details are given in Table 2 and S1.†

6 Conclusions and outlook

In this work, we have generalized '*non-spherical structure refinement*' so that any flavor of quantum-crystallographic crystal-structure treatment can be linked with the established and modern free software *Olex2*, and specifically with its comprehensive refinement engine *olex2.refine*. The capabilities of *NoSpherA2* are demonstrated here for the example of Hirshfeld Atom Refinement, which can now be applied to disordered structures, inorganic periodic network compounds including salts, and compounds containing heavy elements. These developments also open HAR to the fields of protein crystallography, as well as inorganic and metal–organic materials. In this respect, as a core chemical result, we demonstrate that there is a strong directional dependence of bonding and nonsphericity of electron density in fluorite CaF_2, although it is generally assumed that it consists of spherical ions. We will investigate this point further by applying non-spherical refinements to other ionic species.

HARs in *NoSpherA2* are significantly faster than previous implementations of HAR and often more accurate. In principle, any modern quantum-mechanical software can now be used for the theoretical steps, demonstrated here by the use of *ORCA*, which is a QM software freely available for academic use. The PBE DFT method with triple-zeta basis sets is recommended for a good balance between accuracy and speed. This is even true for heavy transition metals, where relativistic extensions are necessary. *NoSpherA2* also enabled us to pin down the problems in the crystallographic refinement of heavy elements and show pathways for further developments in the field.

Since the format file underlying *NoSpherA2* is entirely general, interfacing *Olex2* with other flavors of X-ray non-spherical structure refinement is simple and straightforward. First tests have been made for the HAR-ELMO method[39] and the multipole databank software *Discamb.*[29] Moreover, any of these kinds of non-spherical structure refinement are not restricted to X-radiation. First tests towards the non-spherical refinement of electron-diffraction data have been made inside and outside[107] *NoSpherA2*.

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

The authors acknowledge much feedback of testers and users of *NoSpherA2* alpha versions leading to optimization of the GUI and code and many bug-fixes. We thank Professor Hans-Beat Bürgi for continuous constructive criticism during the development of *NoSpherA2*, and Dr Georgia Cametti for the crystal of CaF_2. We gratefully acknowledge the support for LM through the Intensive Industrial Innovation Programme run by Durham University and funded by the European Regional Development Fund (ERDF).

Notes and references

‡ Three main theories can each be used for both strategies: multipole modelling,[17,18] maximum entropy methods[19] and X-ray wavefunction refinement.[20–22] The latter is part of the original definition of quantum crystallography (QCr),[23] whereas all theories are discussed in the generalized field of QCr.[24]

1 M. Jaskolski, Z. Dauter and A. Wlodawer, *FEBS J.*, 2014, **281**, 3985–4009.

2 J. D. Watson and F. H. C. Crick, *Nature*, 1953, **171**, 737–738.

3 A. K. Geim and K. S. Novoselov, *Nat. Mater.*, 2007, **6**, 183–191.

4 E. Cartlidge, *Science*, 2016, **354**, 22–23.

5 T. E. Mason, D. Abernathy, I. Anderson, J. Ankner, T. Egami, G. Ehlers, A. Ekkebus, G. Granroth, M. Hagen, K. Herwig, J. Hodges, C. Hoffmann, C. Horak, L. Horton, F. Klose, J. Larese, A. Mesecar, D. Myles, J. Neuefeind, M. Ohl, C. Tulka, X.-L. Wang and J. Zhao, *Phys. B*, 2006, **385–386**, 955–960.

6 T. Gruene, J. T. C. Wennmacher, C. Zaubitzer, J. J. Holstein, J. Heidler, A. Fecteau-Lefebvre, S. De Carlo, E. Müller, K. N. Goldie, I. Regeni, T. Li, G. Santiso-Quinones, G. Steinfeld, S. Handschin, E. van Genderen, J. A. van Bokhoven, G. H. Clever and R. Pantelic, *Angew. Chem., Int. Ed.*, 2018, **57**, 16313–16317.

7 A. H. Compton, *Nature*, 1915, **95**, 343–344.

8 H. Kasai, K. Tolborg, M. Sist, J. Zhang, V. R. Hathwar, M. Ø. Filsø, S. Cenedese, K. Sugimoto, J. Overgaard, E. Nishibori and B. B. Iversen, *Nat. Mater.*, 2018, **17**, 249–252.

9 H. Keil, M. Hellström, C. Stückl, R. Herbst-Irmer, J. Behler and D. Stalke, *Chem. – Eur. J.*, 2019, **25**, 15786–15794.

10 C. Jelsch, M. M. Teeter, V. Lamzin, V. Pichon-Pesme, R. H. Blessing and C. Lecomte, *Proc. Natl. Acad. Sci. U. S. A.*, 2000, **97**, 3171–3176.

11 M. Fugel, L. A. Malaspina, R. Pal, S. P. Thomas, M. W. Shi, M. A. Spackman, K. Sugimoto and S. Grabowsky, *Chem. – Eur. J.*, 2019, **25**, 6523.

12 M. Woińska, S. Grabowsky, P. M. Dominiak, K. Woźniak and D. Jayatilaka, *Sci. Adv.*, 2016, **2**(5), e1600192.

13 A. A. Hoser and A. Ø. Madsen, *Acta Crystallogr., Sect. A: Found. Adv.*, 2017, **73**, 102–114.

14 M. J. Turner, S. Grabowsky, D. Jayatilaka and M. A. Spackman, *J. Phys. Chem. Lett.*, 2014, **5**, 4249–4255.

15 S. P. Thomas, P. R. Spackman, D. Jayatilaka and M. A. Spackman, *J. Chem. Theory Comput.*, 2018, **14**, 1614–1623.

16 F. Kleemiss, A. Justies, D. Duvinage, P. Watermann, E. Ehrke, K. Sugimoto, M. Fugel, L. A. Malaspina, A. Dittmer, T. Kleemiss, P. Puylaert, N. R. King, A. Staubitz, T. M. Tzschentke, R. Dringen, S. Grabowsky and J. Beckmann, *J. Med. Chem.*, 2020, DOI: 10.1021/acs.jmedchem.0c00813, in press.

17 R. F. Stewart, *J. Chem. Phys.*, 1969, **51**, 4569–4577.

18 N. K. Hansen and P. Coppens, *Acta Crystallogr., Sect. A: Cryst. Phys., Diffr., Theor. Gen. Crystallogr.*, 1978, **34**, 909–921.

19 (*a*) D. M. Collins, *Nature*, 1982, **298**, 49–51; (*b*) M. Sakata and M. Sato, *Acta Crystallogr., Sect. A: Found. Crystallogr.*, 1990, **46**, 263.

20 M. Woińska, D. Jayatilaka, B. Dittrich, R. Flaig, P. Luger, K. Woźniak, P. M. Dominiak and S. Grabowsky, *ChemPhysChem*, 2017, **18**, 3334–3351.

21 C. Gatti and P. Macchi, *Modern Charge-Density Analysis, Springer*, Dordrecht, 2012, ISBN: 978-90-481-3835-7.

22 D. Stalke, Electron Density and Chemical Bonding I: Experimental Charge Density Studies, in *Structure and Bonding*, vol 146, Springer, Dordrecht, 2012.

23 S. Grabowsky, A. Genoni and H.-B. Bürgi, *Chem. Sci.*, 2017, **8**, 4159–4176.

24 A. Genoni, L. Bučinský, N. Claiser, J. Contreras-García, B. Dittrich, P. M. Dominiak, E. Espinosa, C. Gatti, P. Giannozzi, J.-M. Gillet, D. Jayatilaka, P. Macchi, A. Ø. Madsen, L. Massa, C. F. Matta, K. M. Merz Jr, P. N. H. Nakashima, H. Ott, U. Ryde, K. Schwarz, M. Sierka and S. Grabowsky, *Chem. – Eur. J.*, 2018, **24**, 10881–10905.

25 (*a*) D. T. Cromer and J. B. Mann, *Los Alamos Scientific Laboratory report LA-3816*, 1968; (*b*) D. T. Cromer and J. T. Waber, *International tables for X-ray crystallography*, 1974, vol. IV, 71.

26 B. Dittrich, C. B. Hübschle, K. Pröpper, F. Dietrich, T. Stolper and J. J. Holstein, *Acta Crystallogr., Sect. B: Struct. Sci., Cryst. Eng. Mater.*, 2013, **69**, 91–104.

27 K. N. Jarzembska and P. M. Dominiak, *Acta Crystallogr., Sect. A: Found. Crystallogr.*, 2012, **68**, 139–147.

28 S. Domagała, B. Fournier, D. Liebschner, B. Guillot and C. Jelsch, *Acta Crystallogr., Sect. A: Found. Crystallogr.*, 2012, **68**, 337–351.

29 K. K. Jha, B. Gruza, P. Kumar, M. L. Chodkiewicz and P. M. Dominiak, *Acta Crystallogr., Sect. B: Struct. Sci., Cryst. Eng. Mater.*, 2020, **76**, 296–306.

30 J. Lübben, C. M. Wandtke, C. B. Hübschle, M. Ruf, G. M. Sheldrick and B. Dittrich, *Acta Crystallogr., Sect. A: Found. Adv.*, 2019, **75**, 50–62.

31 B. Dittrich and D. Jayatilaka, *Acta Crystallogr., Sect. A: Found. Crystallogr.*, 2008, **64**, 383–393.

32 S. C. Capelli, H.-B. Bürgi, B. Dittrich, S. Grabowsky and D. Jayatilaka, *IUCrJ*, 2014, **1**, 361–379.

33 F. L. Hirshfeld, *Theor. Chim. Acta*, 1977, **44**, 129–138.

34 W. F. Sanjuan-Szklarz, M. Woińska, S. Domagała, P. M. Dominiak, S. Grabowsky, D. Jayatilaka, M. Gutmann and K. Woźniak, *IUCrJ*, 2020, **7**, 920–933.

35 D. Jayatilaka and D. J. Grimwood, in *Computational Science — ICCS* 2003, Springer Berlin Heidelberg, Berlin, Heidelberg, vol. 2003, pp. 142–151.

36 L. A. Malaspina, A. Genoni and S. Grabowsky, *J. Appl. Crystallogr.*, 2020, under revision.

37 M. Fugel, D. Jayatilaka, E. Hupf, J. Overgaard, V. R. Hathwar, P. Macchi, M. J. Turner, J. A. K. Howard, O. V. Dolomanov, H. Puschmann, B. B. Iversen, H.-B. Bürgi and S. Grabowsky, *IUCrJ*, 2018, **5**, 32–44.

38 B. Meyer and A. Genoni, *J. Phys. Chem. A*, 2018, **122**, 8965–8981.

39 L. A. Malaspina, E. K. Wieduwilt, J. Bergmann, F. Kleemiss, B. Meyer, M. F. Ruiz-López, R. Pal, E. Hupf, J. Beckmann, R. O. Piltz, A. J. Edwards, S. Grabowsky and A. Genoni, *J. Phys. Chem. Lett.*, 2019, **10**(22), 6973–6982.

40 M. E. Wall, *IUCrJ*, 2016, **3**(4), 237–246.

41 C. R. Groom, I. J. Bruno, M. P. Lightfoot and S. C. Ward, The Cambridge Structural Database, *Acta Crystallogr., Sect. B: Struct. Sci., Cryst. Eng. Mater.*, 2016, **72**, 171–179, accessed June 18th, 2020.

42 J. J. Lee, A. N. Sobolev, M. J. Turner, R. O. Fuller, B. B. Iversen, G. A. Koutsantonis and M. A. Spackman, *Cryst. Growth Des.*, 2014, **14**, 1296–1306.

43 H. Furukawa, K. E. Cordova, M. O'Keeffe and O. M. Yaghi, *Science*, 2013, **341**, 1230444.

44 P. J. Waller, F. Gándara and O. M. Yaghi, *Acc. Chem. Res.*, 2015, **48**, 3053–3063.

45 M. Hoshino, A. Khutia, H. Xing, Y. Inokuma and M. Fujita, *IUCrJ*, 2016, **3**, 139–151.

46 J.-P. Torré, R. Coupan, M. Chabod, E. Pere, S. Labat, A. Khoukh, R. Brown, J.-M. Sotiropoulos and H. Gornitzka, *Cryst. Growth Des.*, 2016, **16**, 5330–5338.

47 J.-P. Torré, H. Gornitzka, R. Coupan, C. Dicharry, M. Pérez-Rodríguez, A. Comesaña and M. M. Piñeiro, *J. Phys. Chem. C*, 2019, **123**, 14582–14590.

48 C. A. Hargreaves, G. Sandford, R. Slater, D. S. Yufit, J. A. K. Howard and A. Vong, *Tetrahedron*, 2007, **63**, 5204–5211.

49 M. Stachowicz, M. Malinska, J. Parafiniuk and K. Woźniak, *Acta Crystallogr., Sect. B: Struct. Sci., Cryst. Eng. Mater.*, 2017, **73**, 643–653.

50 (*a*) P. Seiler and J. D. Dunitz, *Helv. Chim. Acta*, 1986, **69**, 1107–1112; (*b*) M. G. Trefry, E. N. Maslen and M. A. Spackman, *J. Phys. C: Solid State Phys.*, 1987, **20**, 19–28.

51 M. C. Schmidt, R. Colella and D. R. Yoder-Short, *Acta Crystallogr., Sect. A: Found. Crystallogr.*, 1985, **41**, 171–175.

52 J. M. Zuo, M. O'Keeffe, P. Rez and J. C. H. Spence, *Phys. Rev. Lett.*, 1997, **78**, 4777–4780.

53 I. M. Kurylyshyn, T. F. Fässler, A. Fischer, C. Hauf, G. Eickerling, M. Presnitz and W. Scherer, *Angew. Chem., Int. Ed.*, 2014, **53**, 3029–3032.

54 R. Pal, S. Mebs, M. W. Shi, D. Jayatilaka, J. M. Krzeszczakowska, L. A. Malaspina, M. Wiecko, P. Luger, M. Hesse, Y.-S. Chen, J. Beckmann and S. Grabowsky, *Inorg. Chem.*, 2018, **57**, 4906–4920.

55 R. B. King, *Chem. Rev.*, 2001, **101**, 1119–1152.

56 S. Mebs, R. Kalinowski, S. Grabowsky, D. F. Forster, R. Kickbusch, E. Justus, W. Morgenroth, C. Paulmann, P. Luger, D. Gabel and D. Lentz, *Inorg. Chem.*, 2011, **50**, 90–103.

57 S. Mebs, J. Henn, P. Luger and D. Lentz, *Z. Anorg. Allg. Chem.*, 2013, **639**, 2057–2064.

58 Q. L. Lu, J. W. Meng, W. J. Song and J. G. Wan, *Int. J. Hydrogen Energy*, 2013, **38**, 13328–13334.

59 G. J. Kubas, *Metal Dihydrogen and s-Bond Complexes in Modern Inorganic Chemistry*, Kluwer Academic/Plenum Publishers, New York, 2001.

60 L. Bučinský, D. Jayatilaka and S. Grabowsky, *J. Phys. Chem. A*, 2016, **120**, 6650–6669.

61 C. Gao, A. Genoni, S. Gao, S. Jiang, A. Soncini and J. Overgaard, *Nat. Chem.*, 2020, **12**, 213–219.

62 P. W. Frost, J. A. K. Howard and J. L. Spencer, *Acta Crystallogr., Sect. C: Cryst. Struct. Commun.*, 1984, **40**, 946–949.

63 J. A. K. Howard, O. Johnson, T. F. Koetzle and J. L. Spencer, *Inorg. Chem.*, 1987, **26**, 2930–2933.

64 (*a*) GUI: O. V. Dolomanov, L. J. Bourhis, R. J. Gildea, J. A. K. Howard and H. Puschmann, *J. Appl. Crystallogr.*, 2009, **42**, 339–341; (*b*) Refinement engine olex2.refine: L. J. Bourhis, O. V. Dolomanov, R. J. Gildea, J. A. K. Howard and H. Puschmann, *Acta Crystallogr., Sect. A: Found. Adv.*, 2015, **71**, 59–75.

65 L. Midgley, L. J. Bourhis, O. V. Dolomanov, N. Peyerimhoff and H. Puschmann, 2019, arXiv:1911.08847.

66 L. Midgley, L. J. Bourhis, O. V. Dolomanov, S. Grabowsky, F. Kleemiss, N. Peyerimhoff, H. Puschmann, Nonspherical Structure Refinement (NoSpherA2) in olex2.refine: A Mathematical Analysis, in preparation.

67 F. Neese, *Wiley Interdiscip. Rev.: Comput. Mol. Sci.*, 2012, **2**, 73–78. For Version 4.0 and above also: F. Neese, *Wiley Interdiscip. Rev.: Comput. Mol. Sci.*, 2018, **8**, e1327.

68 M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, G. A. Petersson, H. Nakatsuji, X. Li, M. Caricato, A. Marenich, J. Bloino, B. G. Janesko, R. Gomperts, B. Mennucci, H. P. Hratchian, J. V. Ortiz, A. F. Izmaylov, J. L. Sonnenberg, D. Williams-Young, F. Ding, F. Lipparini, F. Egidi, J. Goings, B. Peng, A. Petrone, T. Henderson, D. Ranasinghe, V. G. Zakrzewski, J. Gao, N. Rega, G. Zheng, W. Liang, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida,

T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, K. Throssell, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, T. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, J. M. Millam, M. Klene, C. Adamo, R. Cammi, J. W. Ochterski, R. L. Martin, K. Morokuma, O. Farkas, J. B. Foresman, and D. J. Fox, *Gaussian 09, Revision D.03*, Gaussian, Inc., Wallingford C. T., 2016.

69 P. Pollak and F. Weigend, *J. Chem. Theory Comput.*, 2017, **13**, 3696–3705.

70 A. Wolf, M. Reiher and B. A. Hess, *J. Chem. Phys.*, 2002, **117**, 9215.

71 R. Bast, Numgrid: Numerical integration grid for molecules, DOI: 10.5281/zenodo.3746461, 2020.

72 R. Lindh, P.-Å. Malmqvist and L. Gagliardi, *Theor. Chem. Acc.*, 2001, **106**, 178–187.

73 L. Lebedev, *Russian Academy of Sciences Doklady Mathematics*, 1999, **59**(3), 477–481.

74 J. Burkhardt, , SPHERE_LEBEDEV_RULE: Quadrature Rules for the Unit Sphere, , 2010, https://people.sc.fsu.edu/~jburkardt/cpp_src/sphere_lebedev_rule/sphere_lebedev_rule.html.

75 A. D. Becke, *J. Chem. Phys.*, 1988, **88**, 2547–2553.

76 For Z=1 to Z=36: T. Koga, K. Kanayama, S. Watanabe and A. J. Thakkar, *Int. J. Quantum Chem.*, 1999, **71**, 491–497. For $Z = 37$ to $Z = 86$: T. Koga, K. Kanayama, T. Watanabe, T. Imai and A. J. Thakkar, *Theor. Chem. Acc.*, 2000, **104**, 411–413.

77 M. Kohout, *Int. J. Quantum Chem.*, 2004, **97**, 651–658.

78 E. R. Johnson, S. Keinan, P. Mori-Sanchez, J. Contreras-Garcia, A. J. Cohen and W. Yang, *J. Am. Chem. Soc.*, 2010, **132**, 6498–6506.

79 N. G. Connelly, J. A. K. Howard, J. L. Spencer and P. K. Woodley, *Dalton Trans.*, 1984, 2003–2009.

80 E. K. Wieduwilt, G. Macetti, L. A. Malaspina, D. Jayatilaka, S. Grabowsky and A. Genoni, *J. Mol. Struct.*, 2020, **1209**, 127934.

81 Y.-J. Lee, K. W. Han, J. S. Jang, T.-I. Jeon, J. Park, T. Kawamura, Y. Yamamoto, T. Sugahara, T. Vogt, J.-W. Lee, Y. Lee and J.-H. Yoon, *ChemPhysChem*, 2011, **12**, 1056–1059.

82 A. Nemkevich, M. A. Spackman and B. Corry, *Chem. – Eur. J.*, 2013, **19**, 2676–2684.

83 J.-W. Lee, J. Poudel, M. Cha, S. J. Yoon and J.-H. Yoon, *Energy Fuels*, 2016, **30**, 7604–7609.

84 M. J. Turner, J. J. McKinnon, S. K. Wolff, D. J. Grimwood, P. R. Spackman, D. Jayatilaka and M. A. Spackman, *CrystalExplorer17*, The University of Western Australia, 2017.

85 M. J. Turner, J. J. McKinnon, D. Jayatilaka and M. A. Spackman, *CrystEngComm*, 2011, **13**, 1804–1813.

86 M. A. Spackman and D. Jayatilaka, *CrystEngComm*, 2009, **11**, 19–32.

87 R. F. W. Bader, *Atoms in Molecules: A Quantum Theory*, Clarendon Press, Oxford, 1994.

88 R. F. W. Bader and M. E. Stephens, *J. Am. Chem. Soc.*, 1975, **97**, 7391–7399.

89 M. Kaupp, *Angew. Chem., Int. Ed.*, 2001, **40**, 3534–3565.

90 M. Fugel, M. F. Hesse, R. Pal, J. Beckmann, D. Jayatilaka, M. J. Turner, A. Karton, P. Bultinck, G. S. Chandler and S. Grabowsky, *Chem. – Eur. J.*, 2018, **24**, 15275–15286.

91 V. G. Tsirelson, A. S. Avilov, G. G. Lepeshov, A. K. Kulygin, J. Stahn, U. Pietsch and J. C. H. Spence, *J. Phys. Chem. B*, 2001, **105**, 5068–5074.

92 Search in the Cambridge Structural Database for ammonium ions determined by single-crystal neutron diffraction (June 22nd, 2020). The value given is an average of 11 entries.

93 M. A. Fox, A. E. Goeta, J. A. K. Howard, A. K. Hughes, A. L. Johnson, D. A. Keen, K. Wade and C. C. Wilson, *Inorg. Chem.*, 2001, **40**(1), 173–175.

94 R. Brill, H. Dietrich and H. Dierks, *Acta Crystallogr., Sect. B: Struct. Crystallogr. Cryst. Chem.*, 1971, **27**, 2003–2018.

95 A. Burkhardt, U. Wedig, H. G. von Schnering and A. Savin, *Z. Anorg. Allg. Chem.*, 1993, **619**, 437–441.

96 C. Börrnert, Y. Grin and F. R. Wagner, *Z. Anorg. Allg. Chem.*, 2013, **639**, 2013–2024.

97 L. Bučinský, D. Jayatilaka and S. Grabowsky, *Acta Crystallogr., Sect. A: Found. Adv.*, 2019, **75**, 705–717.

98 M. Podhorský, L. Bučinský, D. Jayatilaka and S. Grabowsky, *Acta Crystallogr., Sect. A: Found. Adv.*, 2020, accepted.

99 J. Christensen, P. N. Horton, C. S. Bury, J. L. Dickerson, H. Taberman, E. F. Garman and S. J. Coles, *IUCrJ*, 2019, **6**, 703–713.

100 B. Cordero, V. Gómez, A. E. Platero-Prats, M. Reves, J. Echeverría, E. Cremades, F. Barragán and S. Alvarez, *Dalton Trans.*, 2008, **21**, 2832.

101 M. G. Medvedev, I. S. Bushmarinov, J. Sun, J. P. Perdew and K. A. Lyssenko, *Science*, 2017, **355**, 49–52.

102 S. Sasaki, *Report No. KEK–88-14, National Lab. for High Energy Physics*, 1989.

103 B. Niepötter, R. Herbst-Irmer and D. Stalke, *J. Appl. Crystallogr.*, 2015, **48**, 1485–1497.

104 R. Destro, R. E. Marsh and R. Bianchi, *J. Phys. Chem.*, 1988, **92**, 966–973.

105 S. C. Capelli, H.-B. Bürgi, S. A. Mason and D. Jayatilaka, *Acta Crystallogr., Sect. C: Struct. Chem.*, 2014, **70**, 949–952.

106 R. M. Kabbani, *Polyhedron*, 1996, **15**, 1951–1955.

107 B. Gruza, M. L. Chodkiewicz, J. Krzeszczakowska and P. M. Dominiak, *Acta Crystallogr., Sect. A: Found. Adv.*, 2020, **76**, 92–109.

3.2 Averaged Interaction Energies (aIE)

The calculation of averaged interaction energies is based on a bash-script (see Listing B.1 in the appendix) that automatically breaks a series of large *.xyz* files into smaller pieces. It starts from a MD simulation that contains the trajectory of individual fragments, adds hydrogen atoms to saturate the backbone of the amino acids and performs wavefunction calculations on them. Then *Tonto* is called in order to calculate the four energy terms of interaction energies as mentioned in equation 2.92 corresponding to electrostatic, dispersion, polarization and repulsion energy on the basis of the calculated wavefunctions. [396, 436, 437]
This aIE script expects all fragments named in the fragments variable to have a *.xyz* file in the current working directory, the ligand geometry and a compiled *Tonto* executable to be present in the folder, as well as a PBS queuing system, where all jobs will be submitted. This PBS system is required in the presented version, while an offline version was also implemented but not used during the course of this thesis. This offline version circumvents the necessity of a computer cluster, but drastically increases the computational load on the single computer. After the calculations finished, a summary of energies can be obtained using the script given in Listing B.2 in the appendix. It would also be easy to include different software for the QM calculation, e.g. *ORCA* or *pySCF*, which would then require a quick conversion of the output files into the *.fchk* format, as this is the only format *Tonto* can read for this analysis, which would not be problematic using *cuQCrT* or a subset of its functions. This was not done here, since *Gaussian09* was available. [438]

The resulting files *disp.ener*, *ele.ener*, *pol.ener* and *rep.ener* contain a comma separated list of all energies throughout the simulation for each residue given in the fragments variable. It can be conveniently converted in a spreadsheet to perform analysis and visualization later on. A level of theory for the wavefunction calculation of B3LYP/6-31G(d,p) was chosen for this test, reflected by the correpsponding lines in the aIE script as seen in appendix B.1, to match the *CrystalExplorer* settings and reported scaling factors for total interaction energies. [289, 396, 423] Total energies, which are usually only computed by *Crystal Explorer* based on the values read from the *Tonto* output, are obtained by importing the *.ener* files as *.csv* (=comma separated values) into a spreadsheet software and performing the sum of the four different contributions based on the pre-defined scaling factors matching the selected level of theory. In this case of B3LYP/6-31G(d,p) the total energy results as:

$$E_{Tot} = 1.057E_{ele} + 0.74E_{pol} + 0.871E_{dis} + 0.618E_{rep}.$$

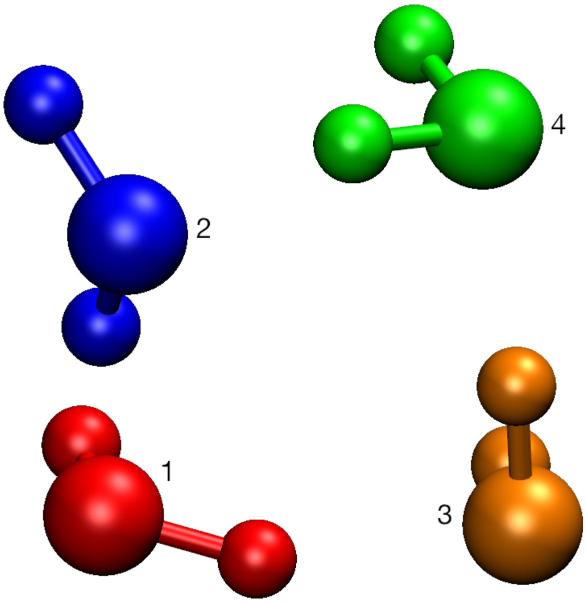
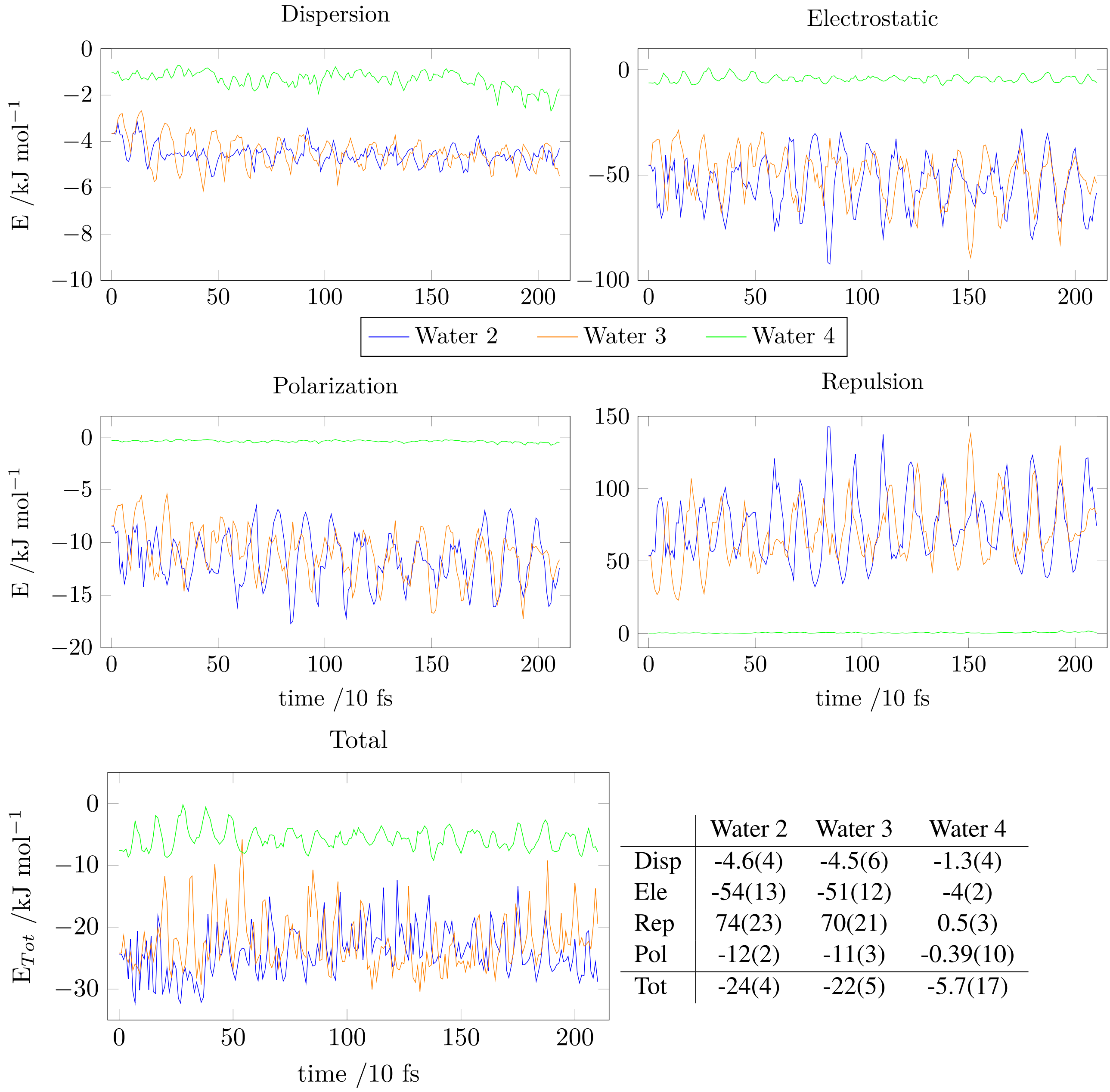


Figure 3.7: Visualization of the four water molecules used for the *ab-initio* MD. Molecules are color coded and labelled with the number used to refer to them. Molecule 1 (red) was used as the reference molecule, for which the interaction energy with all other molecules was evaluated.

To show how the resulting energies might be interpreted an *ab-initio* MD of four water molecules (see Figure 3.7) was performed using *ORCA* on a level of theory of RI-PBE-D3BJ/def2-SVP with a CPCM model for water to simulate at 300 K for 2.1 ps using a time step of 1 fs. The geometry was saved every 10 fs and used as input for the scripts. A cluster of 4 molecules was chosen to force the molecules into unfavoured interactions, since in normal ice water would form 6-membered rings of molecules. This arrangement should introduce enough motion to give an insight into how they will be depicted by the evolution of interaction energies over time and their averaging. The evolution of the different energy contributions is visualized in Figure 3.8.
The geometrical arrangement of the four molecules has a direct correlation with the distribution of the energy terms among the three interaction energy plots. While moelcule 2 and 3 shown significantly higher absolute



	Water 2	Water 3	Water 4
Disp	-4.6(4)	-4.5(6)	-1.3(4)
Ele	-54(13)	-51(12)	-4(2)
Rep	74(23)	70(21)	0.5(3)
Pol	-12(2)	-11(3)	-0.39(10)
Tot	-24(4)	-22(5)	-5.7(17)

Figure 3.8: Plot of interaction energy contributions and total interaction energy for the water example over time and table of averaged contributions and total scaled energy in kJ mol^{-1} with standard deviations.

values of energies in both directions, attractive and repulsive the realtively distant molecule 4 has comparably small absolute values in terms of energy contributions. The dispersion and electrostatic energy for molecule 4 are only approximately 25 % of the corresponding energies between molecule 1 and 2 or 3, respectively. The polarization and repulsion terms for molecule 4 are nearly zero. The same energy terms are very significant for the other two interaction partner.

In all energy plots the energy is never a constant line but highly affected by more or less periodic fluctuations. These fluctuations might have a simple origin: The thermal coupling of the simulation rescales atomic velocities in certain time intervals and therefore forces atoms to move away or closer to each other than the forces acting in the system would dictate. This mechanism is, however, required in such a small simulation system to make sure the temperature stays sufficiently constant throughout the simulation. A less technical explanation of these effects might be the fluctuation of the arrangement of the four molecules. This arrangement forces the

water molecules to interact with a certain amount of strain in the system, since a 4-membered arrangement of water molecules is not energetically favourable. Therefore the big jumps in energies, that occur in many cases simultaneous between the two closest water molecules might be a result of geometrical strain. To elucidate further on this a more extended study needs to be conducted.
The averaged energies over this simulation show that even though major fluctuations and time dependent energetic extremal situations are observed the averaging allows estimation of the overall interaction energy which is identical within deviations between water molecule 1 and molecules 2 and 3. In addition the approximately 75 % lower interaction energy between water molecule 1 and 4 shows that even minor interactions over a longer range can have a significant contribution that is higher than the thermal fluctuation. These findings are a proof of concept for the investigations performed in sections 5.2.10 & 5.2.14.

3.3 The CUDA-accelerated Quantum-Crystallographic-Toolbox (cuQCrT)

The work on differences in effects onto the electron density and potentials from different environments and the evaluation of properties from wavefunctions obtained by different programs, especially *Tonto* for XWR, a software was written that could handle these various tasks. Its name is the acronym for the **CU**DA-accelerated **Q**uantum-**Cr**ystallographic-**T**oolbox (cuQCrT). This software was written from scratch and only parts referring to the evaluation of wavefunctions were referenced against *Tonto*, *Multiwfn* or *nciplot*. [153, 439, 440] The main applications can be summarized in the following points:

1. Conversion of *.wfn* or *.wfx* files into *.fchk* files (based on input from Alessandro Genoni on the syntax of the *.fchk* file)
2. Calculation of properties based on a *.wfn/.wfx* file for a complete molecule
3. Calculation of averaged properties based on MD or QM/MM trajectories
4. Calculation of bond- or atom-centered and scaled areas of properties from a *.wfn/.wfx* file
5. Working with *.cube* files (manipulation, analysis, comparison etc. based on input from Emmanuel Hupf on the initial drafts of *.cube* comparison routines)

These features, especially making the wavefunction obtained by XWR using *Tonto* available to the much more widespread *Gaussian* file format *.fchk*, were a crucial step towards the development of Complementary X-ray constrained Bonding Analysis, as introduced in section 1.3 and allowed the fast implementation of *NoSpherA2* using the existing framework for wavefunction handling and property calculations.
The question whether there is a need for another software (compare DGrid, Multiwfn, NCIplot etc. [439–441]) capable of file conversion and calculation of properties might arise. The major advantage of *cuQCrT* is the *CUDA* acceleration. This is a framework for programming in different languages to make use of the GPU, which is a highly parallel computational setup that can perform many calculations simultaneously in a parallel manner on graphics cards. [442] This allows the evaluation of properties from a wavefunction, to give one example, to be done in a much shorter time compared to the best optimized CPU code on the newest model, due to the sheer amount of processing units present on consumer GPUs. While the evaluation of scalar fields based on a single geometry like the reduced density gradient based on a complete protein like the once presented in section 5.2.14 takes approximately 8 hours on a quad-core CPU the calculation of 1000 different scalar fields using the same input geometries and averaging of the fields took approx. 20 hours on a better consumer GPU (GeForce 1080Ti).

3.3.1 Preparation and Conversion of Wavefunction Files

The conversion of a *.wfn* file into a *.fchk* file is a key step to perform complementary bonding analysis (CBA), as this file format is read by most programs or can easily be transferred into one. Prof. Alessandro Genoni provided a *Fortran* code, that could transfer the MO coefficients into the format and order needed in a *.fchk* file

and calculate the density matrix that is also needed, which can be manually inserted into an existing *.fchk* file. However, the *.wfn* usually only contains occupied orbitals. This would be compensated by putting 0 for all other MO coefficients of virtual orbitals. But especially for NBO analysis this is not useful, as virtual orbitals are mixed with occupied ones. Therefore a modified version of the *.wfn* format was introduced as output from *Tonto*, which also contains the virtual orbitals: the *.ffn*. The code from Prof. Genoni was modernized and adapted in *cuQCrT* to write a full *.fchk* file fully automatically.
The first version of an implementation managed this by running a *Gaussian09* [438] job locally or using an existing *.fchk* file and inserting the calculated information. The current implementation provides the complete file independently, filling in all missing information like atomic masses, basis function information taken from a *Tonto* basis-sets folder and also sorting the MO coefficients of other quantum mechanical software, as each code has a different order for primitives of higher angular momentum.
The resulting *.fchk* can be used e.g. to perform NBO-, [189–191] QTAIM- [177, 178] or ELI-analysis. [186, 187] Bsically all methods, except EDA, which are mentioned in section 1.3 are available using this file. For QTAIM- and ELI-software [441, 443, 444] the *.wfn* or *.wfx* files are in principle suitable, but NBO [445] requires its own input file, which can be generated using *Gaussian* routines which are shipped with the software. First the *unfchk* routine is called to transfer the formatted human-readable file into the binary version readable by the code and then the script shown in listing 3.1 can be used to generate the *.47* file which is the wavefunction archive used by *NBO*-software. It is planned to write conversion routines to directly write the *.47* file using *cuQCrT*, but so far this is not yet implemented. It has to be emphasized that only using the *.ffn* file will work with *NBO*-software, as the unoccupied orbitals are needed to form linear combinations of orbitals.

Listing 3.1: Gaussian Input file for reading in a wavefunction and performing NBO.

```
%chk=./FILENAME.chk
#HF/ChkBasis geom=allcheck guess=(read,only) density=checkpoint pop=nbo6read
    ↪ nosymm 6d 10f

$nbo archive file=OUTPUTNAME $end
```

3.3.2 Real Space Properties

The analysis of real space properties and their partitioning into atomic contributions, basins or bond regions based on topological analysis is not only appealing due to the description in three dimensional space, but also allows intuitive visualization of phenomena. It is common to use meshes or grids of points to evaluate the value of the scalar fields of these properties and use these to draw surfaces of equivalent values or to analyze their spatial distribution based on topological analysis of these discrete grids.
An algorithm to perform topological analysis of a scalar field was implemented in *cuQCrT* based on an implementation of a literature procedure by Tobias Borrmann. [446] The implementation was extended by me to accommodate any size of scalar fields and to work automatized in order to associate the basins found with atoms inside them in case of the scalar field of the electron density. It allows individual print-out of each basin as selected and could be extended to perform a basic implementation of QTAIM analysis.
The calculation of grid based data is evolving around the calculation of ϕ_a and its derivatives at each point in space in a iso-distant point sample. This problem can be iterated in nested loops. This is the case since firstly, the grid properties are completely independent of each other and, secondly, each MO ϕ_a is a sum of AO contributions (compare equation 2.32). The values of ϕ_a are calculated in a loop iterating over all atom primitives χ_μ, calculating their value and the first and second derivatives, as a vector of size 10 (1 for its value, 3 for first derivatives in each spacial direction, 6 for second derivatives including mixed derivatives) and summing them into the corresponding values and derivatives of ϕ_a. Afterwards, density ρ, gradients $\vec{\bigtriangledown}\rho$ and Hessian $\triangle\rho$ at each point in the grid (i, j, k) are calculated as a sum over all MOs using the ath MO occupation number n_a,

dropping indices of grid point for clarity, by:

$$\rho = \sum_a n_a \, \phi_a^2 \tag{3.2}$$

$$\vec{\nabla}\rho = \sum_a 2n_a \begin{pmatrix} \phi_a \frac{\partial \phi_a}{\partial x} \\ \phi_a \frac{\partial \phi_a}{\partial y} \\ \phi_a \frac{\partial \phi_a}{\partial z} \end{pmatrix} \tag{3.3}$$

$$\triangle\rho = \sum_a 2n_a \begin{pmatrix} \phi_a \frac{\partial^2 \phi_a}{\partial x^2} + \left(\frac{\partial \phi_a}{\partial x}\right)^2 & \phi_a \frac{\partial^2 \phi_a}{\partial x \partial y} + \frac{\partial \phi_a}{\partial x}\frac{\partial \phi_a}{\partial y} & \phi_a \frac{\partial^2 \phi_a}{\partial x \partial z} + \frac{\partial \phi_a}{\partial x}\frac{\partial \phi_a}{\partial z} \\ \phi_a \frac{\partial^2 \phi_a}{\partial x \partial y} + \frac{\partial \phi_a}{\partial x}\frac{\partial \phi_a}{\partial y} & \phi_a \frac{\partial^2 \phi_a}{\partial y^2} + \left(\frac{\partial \phi_a}{\partial y}\right)^2 & \phi_a \frac{\partial^2 \phi_a}{\partial y \partial z} + \frac{\partial \phi_a}{\partial y}\frac{\partial \phi_a}{\partial z} \\ \phi_a \frac{\partial^2 \phi_a}{\partial x \partial z} + \frac{\partial \phi_a}{\partial x}\frac{\partial \phi_a}{\partial z} & \phi_a \frac{\partial^2 \phi_a}{\partial y \partial z} + \frac{\partial \phi_a}{\partial y}\frac{\partial \phi_a}{\partial z} & \phi_a \frac{\partial^2 \phi_a}{\partial z^2} + \left(\frac{\partial \phi_a}{\partial z}\right)^2 \end{pmatrix} \tag{3.4}$$

The position in space necessary to calculate these values for each grid point is given by the index of the gridpoint (i, j, k), the origin of the cube $\vec{r}_0$ and the 3x3 matrix of unit vectors $\mathbf{M}$, defining the orientation, shape and grid spacing of the cube, in the following manner:

$$\vec{r} = \vec{r}_0 + \mathbf{M} \cdot \begin{pmatrix} i \\ j \\ k \end{pmatrix} \tag{3.5}$$

The nature of these calculations allows various implementations of parallelism. In *cuQCrT*, the acceleration of GPUs is achieved employing a 2D-mapping of the problem onto the multiprocessor units of the GPU, by spreading the outer loops of grid indices i and j as threads and blocks, respectively, and calculating the last loop over k of grid points in a pillar along the third axis within this thread. This is beneficial, even if it requires transport of the complete information of the wavefunction to the GPU. The clock frequency of GPUs is significantly lower to that of CPUs, but since the average consumer GPU has between 640 to 3500 cores which can calculate independent tasks making the gain in speed overcompensate the slower clock-speed manifold. Although the CUDA code is only applicable to graphics cards of the manufacturer NVIDIA, the type of task separation is also achievable by other implementations of GPU-usage. The OpenCL language should be mentioned in this context, as is allows the shipping of code for various brands of architecture and also for heterogeneous architecture, that is mixed use of CPU and GPU resources without the necessity to know at compile time which hardware is present. [447] This allows the generation of more widely applicable code, but requires more detailed programming and is more difficult to learn, which is why CUDA implementation was done in this work. A basic layout of these calculations calculating e.g. all mentioned derivatives of the electron density would look like the *pseudo-code* Algorithm 3. The required reading of wavefunction files, organization in data structures, estimation of cube size and orientation, the evaluation of option selection and menu structure, as well as output of results were skipped for clarity.

Electron Localization

The computation of Electron Localization and Localizability, accessible through the calculation of ELF and ELI-D, as defined in equations 2.90 and 2.91, can then be carried out easily, since all necessary values of gradients and density are accessible.

Reduced Density Gradient

The reduced density gradient and the signed electron density, needed for the NCI, are calculated according to equation 2.81. The eigenvalues are obtained by the Jacobi Eigenvalue method, [448] as implemented in C++ by John Burkardt.

Algorithm 3 Pseudocode for calculation of grid based density properties of ρ. i,j,k are indices of calculated grid point. r is the coordinates in real space for this grid point, r_0 is the position of point indexed 0,0,0. **M** is the matrix of unit vectors for the grid, μ a primitive of a wavefunction used for the calculation. r_μ is the position of atom with primitive μ, c_μ is the coefficient of this primitive and n_a is the occupation of MO a. $\vec{\nabla}\vec{\nabla}^T$ is the notation for a 3x3 matrix containing all second derivatives in three dimensional space.

```
for i ← 0, $i_{max}$ do
    for j ← 0, $j_{max}$ do
        for k ← 0, $k_{max}$ do
            r ← $r_0$ + M · (i,j,k)
            for all $\mu \leftarrow 0, \mu_{max}$ do
                d ← r − $r_\mu$
                $\chi$ ← value of GTO/STO
                $\vec{\nabla}\chi$ ← derivatives in x,y,z of $\chi$
                $\vec{\nabla}\vec{\nabla}^T \chi$ ← [3][3] second derivatives in x,y,z of $\chi$
                for all a ← 0, $n_{MO}$ do
                    $\phi_a$ += $c_\mu$ * $\chi$
                    $\vec{\nabla}\phi_a$ += $c_\mu$ * $\vec{\nabla}\chi$
                    $\vec{\nabla}\vec{\nabla}^T \phi_a$ += $c_\mu$ * $\vec{\nabla}\vec{\nabla}^T \chi$
                end for
            end for
            for all a ← 0, $n_{MO}$ do
                $\rho$(i,j,k) += $n_a$ * pow($\phi_a$,2)
                $\vec{\nabla}\rho$(i,j,k) += 2 * $n_a$ * $\phi_a$ * $\vec{\nabla}\phi_a$
                $\triangle\rho$(i,j,k) += 2 * $n_a$ * $\phi_a$ * $(\vec{\nabla}\phi_a)^T$ * $\vec{\nabla}\vec{\nabla}^T \phi_a$ * $\vec{\nabla}\phi_a$
            end for
        end for
    end for
end for
```

Total Electrostatic Potential

The calculation of the ESP is separable into two problems, one being rather simple, namely the calculation of nuclear contributions to the total potential, the second one being rather complicated, since it is highly demanding in terms of computational power, as the integral over the electron density requires a double loop over the AOs, which scales to a upper-triangular square of the number of AOs.

3.3.3 Bond Based Calculation of Properties

Due to the ambiguity of the definition of a chemical bond in the context of wavefunctions, as mentioned in the introduction, but the impressive success of the concept of chemical bonding, the analysis of wavefunctions in terms of localized interactions resembling bonding was suggested based on complementary bonding analysis. [172–175, 392, 424]

In this thesis, the effect of environments onto the molecular properties was investigated using the bondwise difference of the electron density or electrostatic potential. These quantities were calculated by *cuQCrT* on a grid that uses the vectors between three atom positions in Cartesian space to define origin and orientation of the cube to calculate. The size of the region to calculate is either selected in absolute length or in multiples of the bond length (modulus of the vector between atom 1 and 2). The number of gridpoints is either specified explicitly (e.g. 200 gridpoints along the x-axis) or selected by a resolution (e.g. 0.01 Å distance between gridpoints). This idea of scaling of boxes was introduced during the master thesis "*Crystal and enzyme environmental effects on the electron density of a cysteine protease inhibitor model compound*". But in this thesis it is implemented in *cuQCrT*, which can calculate these values independently of external software, based on any *.wfn/.wfx*-file. The orientation of the cube among the three atoms is visualized in Figure 3.9.

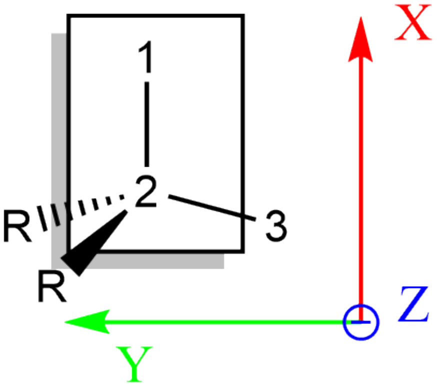


Figure 3.9: Visualization of the positioning of the cube wherein the grid of real space properties is calculated. The local coordinate system is shown in relation to atoms 1,2,3, the box represents the front view of the cube to be considered.

3.3.4 Orbital Space Analysis

The analysis in orbital space, which is the basis for, e.g., NBO (see section 2.3.7) or RGBI (see section 2.3.8), allows understanding of bonding in terminology of bond orbitals. The framework of NBO, namely Natural Population Analysis (NPA), Natural Localized Molecular Orbitals (NLMO) [189] and NRT [449] was analyzed using *NBO* Version 6.0. [445] The resulting wavefunction of XWR is used as input, translated by *cuQCrT* into a formatted checkpoint (*.fchk*)-file (see section 3.3.1). Then the standalone version of *NBO* can be used.
The calculation of RGBI was performed using *Tonto*, reading the wavefunction either directly from *Tonto*-internal file formats or a *.fchk*-file, obtained from *cuQCrT* reading the *.wfn*-file.

3.3.5 X-ray Constrained Complementary Bonding Analysis

The calculation of properties and bonding indicators from wavefunctions provides deep insight into the chemistry of a molecule. However, they are limited by the theoretical model implemented. The most common type of calculations is performed *in vacuo*, that means without any interaction with the environment. This would coincide with the simulated entity being the only particle in existence. This major assumption is sometimes tried to overcome by a simulation of the environment by either explicit or implicit models of the environment. If a situation in solvation is modeled, usually a conductor like model is used [450] to calculate the field and polarization experienced by the electrons in the wavefunction based on a continuous dielectric constant that should mimic the effect of a solvent. This approach is computationally not very costly, but also neglects any explicit interaction like hydrogen bonds or dispersion contribution of neighboring molecules.
A different approach is to explicitly include solvent molecules, which can then describe the interaction with the molecule under investigation much more precisely, but at the cost of many more basis functions to be evaluated in the SCF procedure, which is computationally very costly. The QM/MM approach can provide a mixture of both: The molecule experiences directed charges and dispersion forces through the description of the force field, while the number of basis functions remains unchanged.
For the description of crystalline matter, the situation is more complicated. Since the forces acting inside the crystal are of much higher magnitude than in other environments, [451, 452] the description requires more sophisticated approaches. Periodic boundary conditions require a modification of the mathematical framework to include these periodic constraints in the basis sets. But if XWR is used to derive a wavefunction, which includes information from measured structure factors of a diffraction experiment of a single crystal, the resulting calculated properties from this wavefunction might include additional information from the experimental data that were not introduced with the quantum mechanical method. This method might give insight into effects

that would otherwise be extremely costly in terms of computational effort. One example is electron correlation, usually only correctly modeled using high level post-HF methods, while the information is included in the electron density of the crystal (see section 2.2.3.3). A different effect that is not easy to model with conventional methods is the periodic behavior and polarization on the density of the wavefunction using non-periodic basis functions and methods. Only computationally costly (pseudo-)periodic calculations can model these *a priori*, while XWR might be able to extract them from experimental data, as will be shown later for both EC and polarization (see section 4.1.4).

For X-ray constrained CBA it is crucial to have access to the calculated wavefunctions from XWR. Here *cuQCrT* builds a bridge and makes these investigations possible employing all methods discussed in section 1.3. Their application will be shown in the following chapter 4.

Chapter 4

Applying Quantum Crystallography for Complementary Bonding Analysis

4.1 Experimental Electron Correlation, Exchange Correlation Potential and Polarization Obtained from X-ray Wavefunction Refinement

This section is based on an unpublished draft of a publication called "Experimental Electron Correlation and Polarization Obtained from X-ray Wavefunction Refinement". The first author of this manuscript, Dr. Emanuel Hupf, gave his permission to use figures and discussions from the paper draft, which were partially generated using my code or by Dr. Hupf and PD Dr. Grabowsky. The discussion of this manuscript predates many parts of my work and is highly relevant for the analysis reported in this thesis since it allows important conclusions about XCW.

The purpose of this study is to elucidate the possibility and extent of recoverable physical effects contained in the structure factors using XCW. Many effects are interfering inside a crystal, and this affects the electron density distribution, which therefore can not be correctly described by a pure HF-*ansatz* for the calculation of structure factors. Two major components are Electron Correlation and Polarization. These two effects can, however, be quite well modelled by sophisticated theoretical models and therefore their magnitude and spread in space understood from a theoretical point of view. On the downside, these sophisticated calculations are very time consuming and still can only model their respective assumptions. Whether these effects are obtainable from the structure factors of a crystal lattice has already been discussed, [453–457] but these studies only investigated one of the effects and were only taking into account the theoretically calculated structure factors for the case of electron correlation. Here we show that the experimental data can yield similar effects as the theoretical data and that the separation of effects is possible based on a selection of assumptions in the used *ansatz*. The study can be separated into two parts:

(i) Theoretical wavefunctions were computed using a quantum mechanical *ansatz* that contains a certain effect (Correlated Hamiltonian or polarized environment through periodic boundary conditions during the calculation) and subsequently a pure HF-XCW was performed to retrieve the effect from a set of calculated structure factors

(ii) experimental data was used with a combination of different effects included in the quantum mechanical *ansatz* to model the remaining effects and to check whether the correct effect can be reconstructed. All calculations of this section were performed using the pob-TZVP basis set, which is part of the *CRYSTAL* software package. [458]

Two XCW fittings for each effect were performed for each resolution using different methods to investigate the possibility of XCW to cope with the differences: the first calculation was performed using a pure HF method, which will not contain any effect of electron correlation or polarization. The second calculation was performed using an approximated description of each effect. For electron correlation this calculation is based on the GGA DFT-functional BLYP (compare section 2.2.3.4). This functional has an explicit term to account for electron correlation. Polarization is approximated by the use of an explicit cluster of point charges and dipoles based on

the Hirshfeld stockholder partitioning. These are calculated from an initially calculated wavefunction and will polarize this molecular wavefunction in a field corresponding to all neighbouring Hirshfeld atoms which are generated by the symmetry of the crystal within a certain radius. This makes it an iterative procedure, which will adapt to changes in the wavefunction. In this case, a radius of 8 Å was selected as the cut-off for these charges. My specific contributions to this project were the implementation, calculation and plotting of the exchange correlation potentials on grids inside the program *Tonto* (Figures 4.6-4.8), the final implementation of R$_{RS}$ values and numbers of electron shifted based on an initial script from Dr. Emanuel Hupf, which was extended to correct for double electron density shifts inside my software *cuQCrT*. Additionally, I implemented a procedure to generate bigger grids from a unit cell density grid file through the application of translation vectors, inside *cuQCrT* (see section 3.3), to be able to account for full molecules when only partial molecules are present in the unit cell. This was the case for plots of urea in the periodic boundary calculations using *CRYSTAL14*. For the creation of Figures and calculation of R$_{RS}$ of section 4.1.3 this method was used. To be able to then only analyse a single molecule I implemented a version of the topological analysis of electron density according to *Henkelman et al.* [446] into *cuQCrT*, as described in section 3.3. This algorithm was first implemented for limited cube sizes at compilation time by Dr. Tobias Bormann and extended inside *cuQCrT* to accommodate any sized grid data by using the vector class in C++. I applied it onto unit cell densities to calculate molecular visualizations for pictures including urea from *CRYSTAL14* calculations shown in Figures 4.10 to 4.12.
Finally, I performed data analysis and plots in Figures 4.6 to 4.8 from the XCWs. All other Figures were used with the permission of Dr. Hupf. All the text of this section, as well as the description of Figures, data and results, the discussion and conclusions were written by me. The original draft shows results for two molecules under investigation, urea and L-alanine. Experimental datasets of these molecules were taken from the literature. [434, 459] But since the results show the same trends, only urea will be discussed in this thesis.
Structure factors up to a resolution of $sin\ \theta/\lambda = 2.0$ Å^{-1} were calculated and cut down into two subsets of 1.4415 Å^{-1} and 0.7 Å^{-1}, respectively where 1.4415 Å^{-1} is chosen in agreement with the available resolution of the experimentally available data. [459] Since the calculated theoretical structure factors are in principle without error, but the calculation of χ^2 is dependent on an error assigned to the reflection, these datasets were given a fixed value of 0.1 as the uncertainty.

4.1.1 Electron Correlation (EC)

One of the major effects on the electron density of a molecule compared to the density obtained by a HF calculation is the correlation of electron movement. This effect is modeled in e.g. DFT or post-HF methods (see sections 2.2.3.4 - 2.2.3.3). The investigation using theoretical structure factors is based on a CCSD wavefunction. The difference in electron density between the non-correlated wavefunction from HF and the highly correlated CCSD wavefunction is shown in Figure 4.1. This density grid shows the complete effect of correlation that is included in the structure factors and will be used as the reference. All later calculations will be compared against this reference, as this is the difference that would coincide to a 100% description of the effect in terms of electron density. This is only possible since the absolute magnitude of the effect is known due to the use of theoretical structure factors which were used for the fitting of the wavefunction.

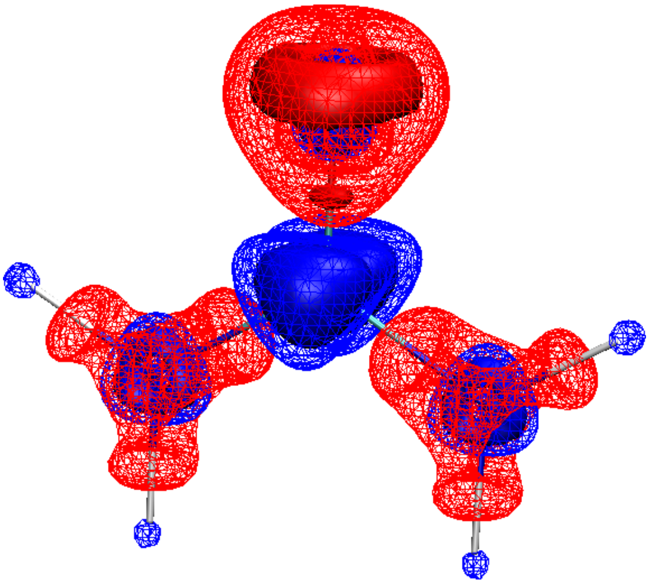


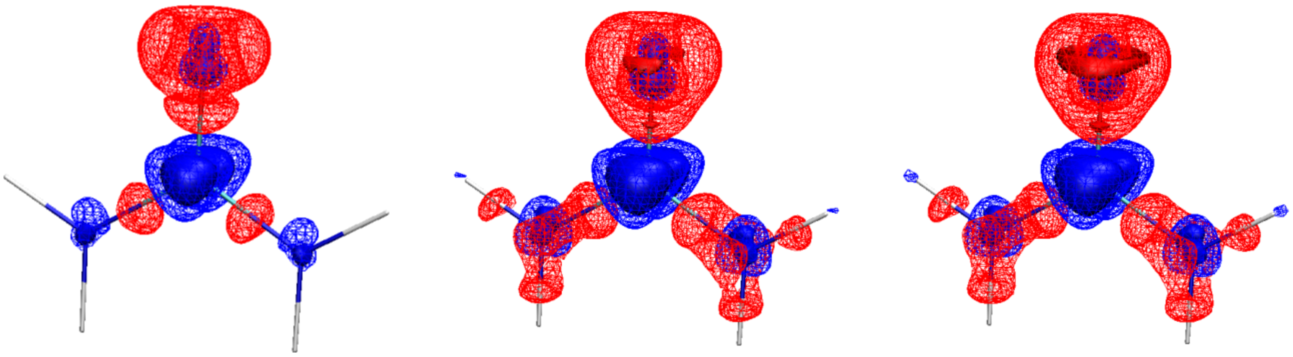
Figure 4.1: Plot of the difference between CCSD and HF wavefunction density used as reference for the reconstruction of XCW fitting. Isosurface values in $e\text{Å}^{-3}$ (solid/wireframe): +0.0025/+0.00125 (blue) and -0.0025/-0.00125 (red).

While the pure HF fitting has to retrieve all effects, the calculation using BLYP-DFT should yield better agreement with the structure factors without any inclusion of structure factor information, as the information is partially included in the *ansatz*. The Figures 4.2 to 4.4 contain the plots for all three resolutions used. The top row (a-c) always refers to the HF based XCW, the bottom row (d-f) to the ones based on BLYP.

The development with increasing lambda at the lowest resolution of 0.7 Å^{-1} shows how the HF wavefunction starts modelling the density better and better with increasing lambda resembling the shape that is expected from the reference density in Figure 4.1. The highest λ value resembles the areas and magnitudes already quite well. In the case of BLYP, one can see an over-expression of the difference at $\lambda = 0$. Since the geometry is kept fixed during XCW what is shown is the pure effect of the functional. Interestingly, the XCW starts compensating the changes, that the functional introduced. The magnitude of difference decreases and the shape becomes closer to the reference density.

The same procedure was repeated for the set of calculated structure factor up to the resolution of the experimentally available dataset in Figure 4.3. It is visually observable that the XCW is going into the same direction as for the lower resolution dataset. At $\lambda = 10.0$ the effect is, however, not completely recovered. The picture looks like it is an intermediate state between the lower resolution fitting snapshots at $\lambda = 0.4$ and 2.0. Also, the compensation of the DFT-XCW seems a bit slower, but not as severely affected as the HF calculation.

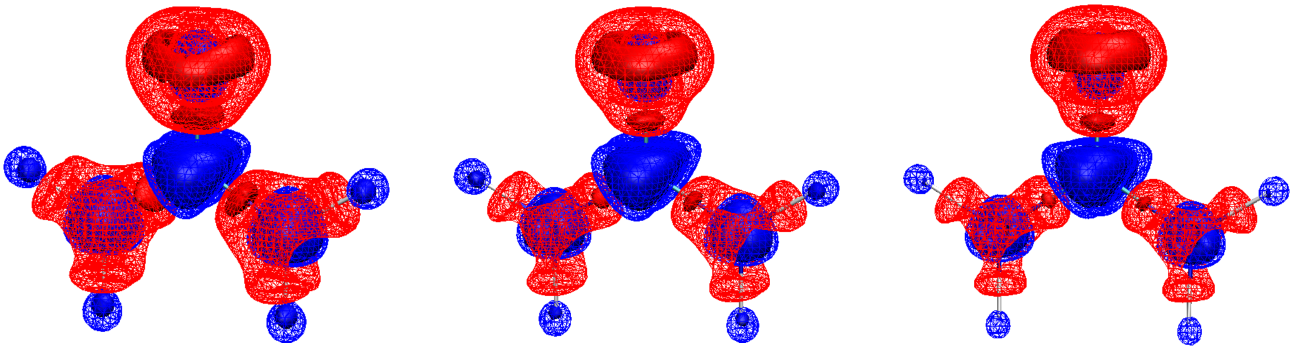
This plot for the highest resolution data is, when looking at the same values of λ, even less advanced in retrieving the effect form the structure factors. The HF-XCW merely has included any effect at $\lambda = 1.0$, while at the lowest resolution already at $\lambda = 0.4$ the picture looked as developed as the one at $\lambda = 10.0$ at this high resolution.



(a) $\lambda_{0.4} - \lambda_{0.0,HF}$ using HF. PR: 70.8%; R_{RS}: 0.0028.

(b) $\lambda_{2.0} - \lambda_{0.0,HF}$ using HF. PR: 90.0%; R_{RS}: 0.0015.

(c) $\lambda_{3.2} - \lambda_{0.0,HF}$ using HF. PR: 94.3%; R_{RS}: 0.0012.



(d) $\lambda_{0.0,BLYP} - \lambda_{0.0,HF}$ using BLYP. PR: 156.5%; R_{RS}: 0.0072.

(e) $\lambda_{0.4,BLYP} - \lambda_{0.0,HF}$ using BLYP. PR: 120.1%; R_{RS}: 0.0031.

(f) $\lambda_{2.0,BLYP} - \lambda_{0.0,HF}$ using BLYP. PR: 109.6%; R_{RS}: 0.0016.

Figure 4.2: Plot of the difference in electron density for different λ values during the XCW using structure factors of a resolution below 0.7 Å^{-1}. Individual details are given in subcaptions. Isosurface values in eÅ^{-3} (solid/wireframe): +0.0025/+0.00125 (blue) and -0.0025/-0.00125 (red).

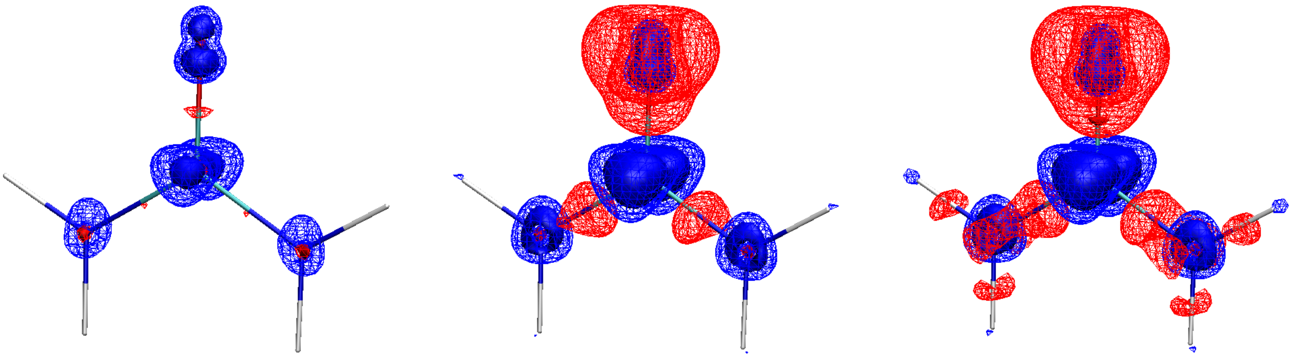
Again, the reconstruction from the exaggerated DFT-XCW does not seem to be as heavily affected, although solid areas in the C-N bonds seem to be more significant at higher λ values.

To understand the differences also in a quantitative matter, the R_{RS} (compare equation 2.80) of the two differences (CCSD-HF vs. XCW-HF/XCW-BLYP) is plotted for all resolutions in Figure 4.5. The Percentage Reconstruction (PR) is constructed as the quotient of N_e (compare equation 2.79) of the reference correlation density plotted in Figure 4.1 and the respective XCW correlation density and also included in Figure 4.5.

The plots of fitting and electron density agreement statistics in Figure 4.5 will quantify the differences throughout the fitting process. The PR was fitted using exponential functions of the form:

$$PR(\lambda) = c + \sum_{i=0}^{3} a_i e^{-b_i \lambda}. \tag{4.1}$$

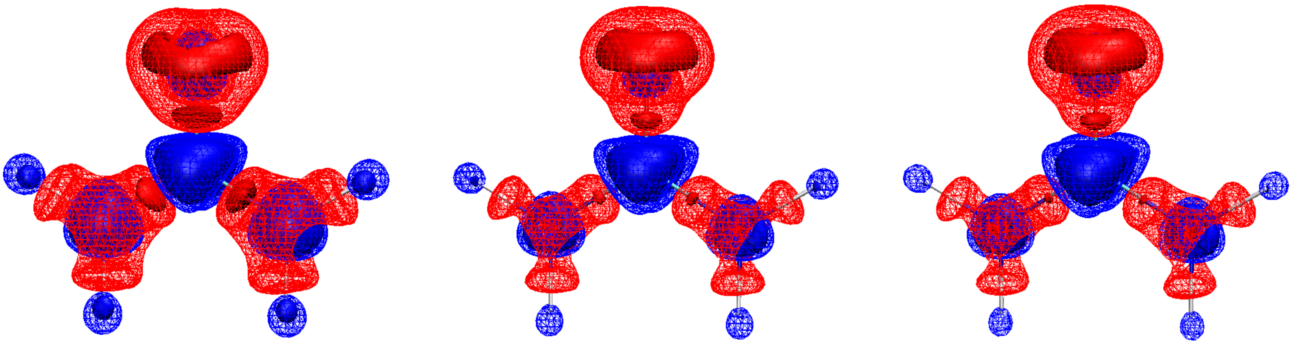
Using these exponential functions the upper/lower limit of the fit and the scale of lambda values needed can be estimated. A table of the fitted values for the exponential functions of the PR (see equation 4.1) is given in table 4.1.



(a) $\lambda_{1.0} - \lambda_{0.0,HF}$ using HF. PR: 44.1%; R$_{RS}$: 0.0049.

(b) $\lambda_{5.0} - \lambda_{0.0,HF}$ using HF. PR: 79.3%; R$_{RS}$: 0.0021.

(c) $\lambda_{10.0} - \lambda_{0.0,HF}$ using HF. PR: 88.4%; R$_{RS}$: 0.0015.

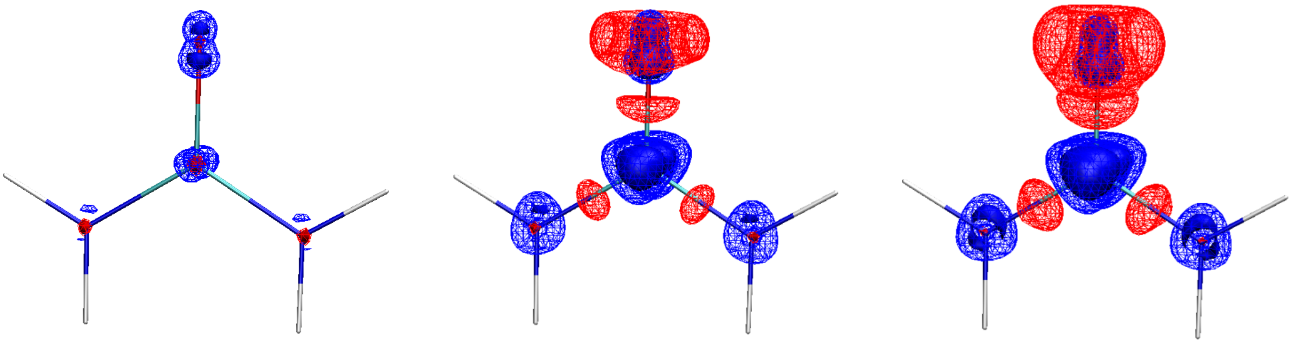


(d) $\lambda_{0.0} - \lambda_{0.0,HF}$ using BLYP. PR: 156.5%; R$_{RS}$: 0.0072.

(e) $\lambda_{5.0} - \lambda_{0.0,HF}$ using BLYP. PR: 113.5%; R$_{RS}$: 0.0021.

(f) $\lambda_{9.5} - \lambda_{0.0,HF}$ using BLYP. PR: 110.2%; R$_{RS}$: 0.0017.

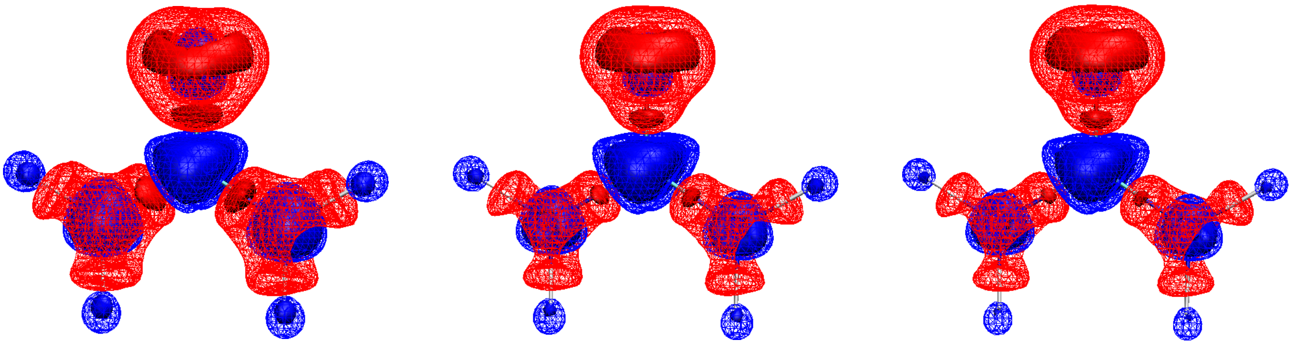
Figure 4.3: Plot of the difference in electron density for different λ values during the XCW using structure factors of a resolution below 1.4415 Å^{-1}. Individual details are given in subcaptions. Isosurface values in eÅ^{-3} (solid/wireframe): +0.0025/+0.00125 (blue) and -0.0025/-0.00125 (red).



(a) $\lambda_{1.0} - \lambda_{0.0,HF}$ using HF. PR: 23.4%; R$_{RS}$: 0.0066.

(b) $\lambda_{5.0} - \lambda_{0.0,HF}$ using HF. PR: 59.4%; R$_{RS}$: 0.0036.

(c) $\lambda_{10.0} - \lambda_{0.0,HF}$ using HF. PR: 74.1%; R$_{RS}$: 0.0024.

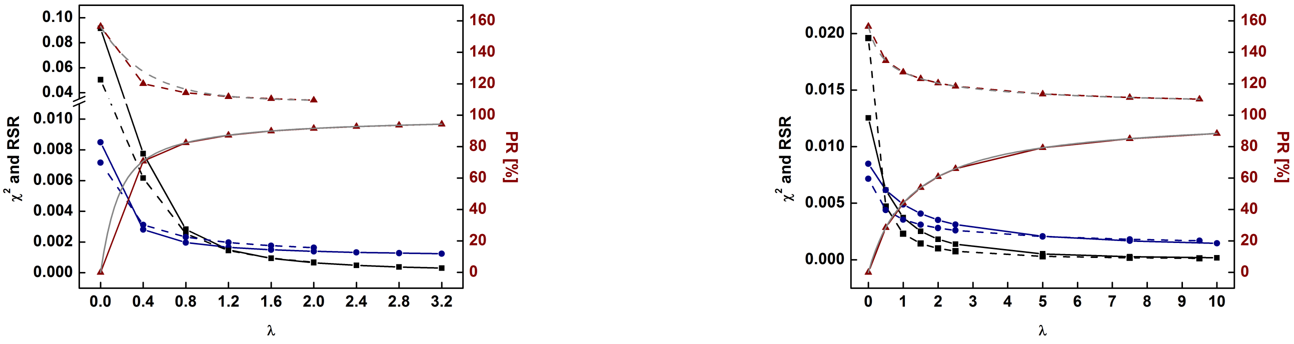


(d) $\lambda_{0.0} - \lambda_{0.0,HF}$ using BLYP. PR: 156.5%; R$_{RS}$: 0.0072.

(e) $\lambda_{5.0} - \lambda_{0.0,HF}$ using BLYP. PR: 120.8%; R$_{RS}$: 0.0028.

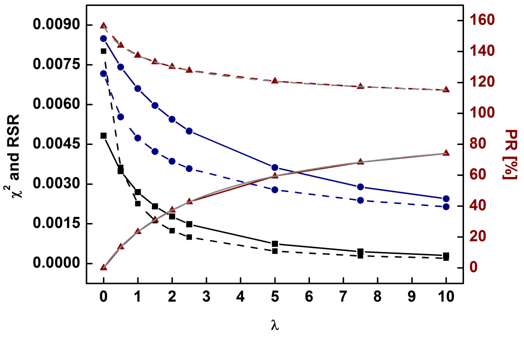
(f) $\lambda_{10.0} - \lambda_{0.0,HF}$ using BLYP. PR: 115.1%; R$_{RS}$: 0.0021.

Figure 4.4: Plot of the difference in electron density for different λ values during the XCW using structure factors of a resolution below 2.0 Å^{-1}. Individual details are given in subcaptions. Isosurface values in eÅ^{-3} (solid/wireframe): +0.0025/+0.00125 (blue) and -0.0025/-0.00125 (red).



(a) (sin $\theta/\lambda)_{max}$ = 0.7 Å^{-1}

(b) (sin $\theta/\lambda)_{max}$ = 1.4415 Å^{-1}



(c) (sin $\theta/\lambda)_{max}$ = 2.0 Å^{-1}

Figure 4.5: Graphs of structure factor and electron density agreement throughout XCWs at all resolutions. χ^2 (black), RSR (=R_{RS}, blue) and the PR (red) plotted against the value of λ during the HF-XCW(solid) or DFT-XCW(dashed). Additionally exponential functions were fitted to the values to estimate the convergence limit of the PR (gray).

Table 4.1: Fitted coefficients of exponential function for reconstructed electron correlation density (see equation 4.1).

resolution /Å^{-1}	method	a_1	b_1	a_2	b_2	a_3	b_3	c
0.7	HF	-12.51	4.47	-30.25	0.57	-57.76	0.14	100.52
1.4415	HF	-34.14	5.49	-42.94	1.30	-16.87	0.42	93.95
2.0	HF	-49.35	8.07	-36.00	1.91	-3.24	0.42	88.59
0.7	BLYP	15.79	0.43	15.79	0.43	15.79	0.43	109.14
1.4415	BLYP	15.00	4.74	19.47	0.96	13.84	0.22	108.18
2.0	BLYP	20.78	6.80	17.92	1.42	7.48	0.33	110.31

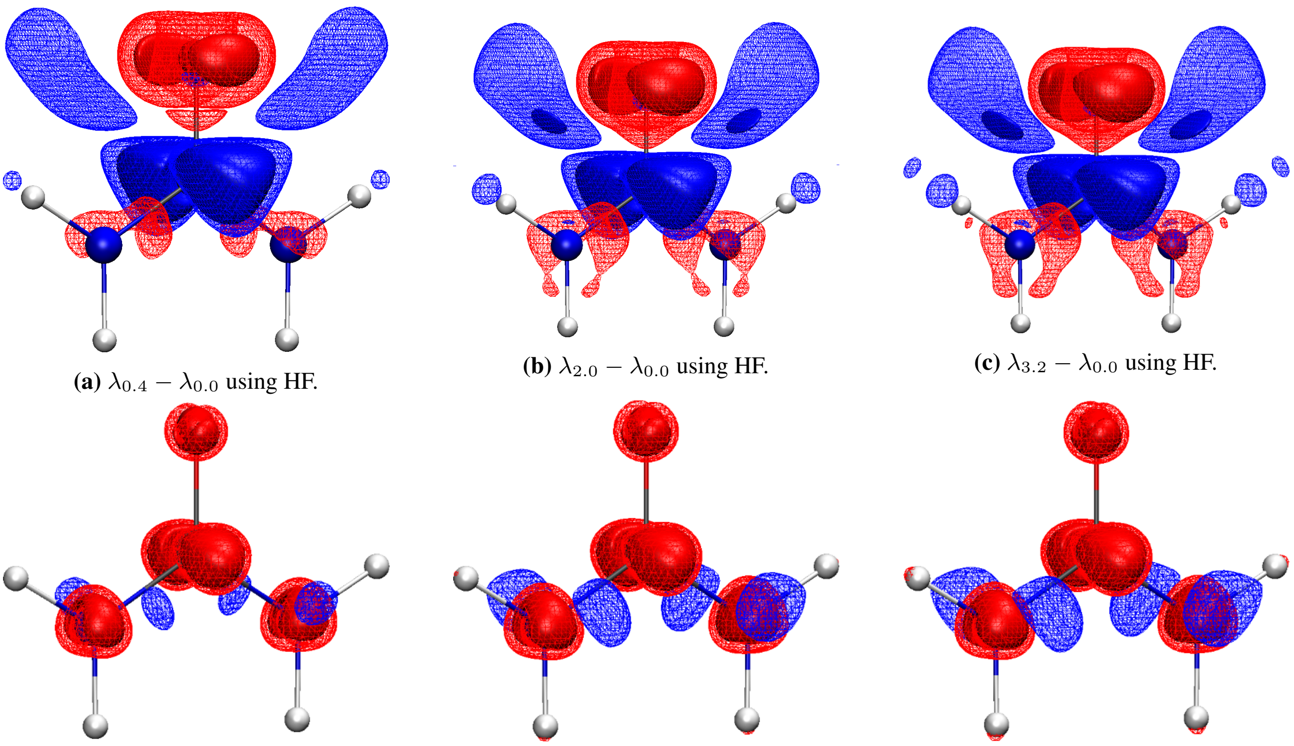
With increased resolution it becomes observable that the leading exponents b_i of all fits become bigger. This corresponds to a slower convergence with respect to increasing λ. To understand why this is the case a look at the values of χ^2 for each resolution might provide the answer. At the higher resolution datasets, the value of χ^2 is drastically decreasing compared to the lower resolution (see Figure 4.5). While χ^2 had a value around 0.091 for the calculation at 0.7 Å^{-1} for λ = 0, the corresponding value for the highest resolution of 2.0 Å^{-1} is 0.0048. This is a drop of one and a half order of magnitudes. But χ^2 is the measure used to introduce the structure factors into the perturbation equation (see section 2.1.9). This means that a much higher value of λ would be required to achieve the same numerical amount of perturbation of the wavefunction through the second term in equation 2.24 during the fitting.

It should be considered that the structure factor difference between correlated and uncorrelated density mainly rely on the low resolution region (compare [455]). Usually, the agreement between even significantly wrong models like the IAM and the experimental density is quite good for high order reflections (compare section 3.1.3). The lower values of χ^2 can be understood as the introduction of many more agreeing reflections with increased resolution than disagreeing reflections are introduced. This is the case since with higher resolutions the number of reflections rises cubically, as the Ewald sphere expands in three dimensions (see section 2.1.3). Interestingly the limit of reconstruction of the effect with $\lim_{\lambda\to\infty} = c$ seems to get worse with increasing resolution. This might be an artifact of the limited λ range used for the fit, but could also be the effect of the introduction of more reflections which decrease the significance of individual disagreeing reflections and therefore lead to incomplete fitting. With increasing resolution the absolute and relative differences of the newly introduced information is smaller, as higher resolution reflections in general contain less information further away from the atom core.

4.1.2 Exchange Correlation Potential

Since the reconstruction of the electron density from structure factors seems to be successful, it is of interest whether XCW can also obtain the correct exchange correlation potential with respect to the reference calculation, which is the key property for accurate description of wavefunctions using DFT. This would allow to fit the Exchange-Correlation potential V_{XC} (see section 2.2.3.4) using crystallographic data from experiments. The exchange correlation potential is a scalar field in 3 dimensional space. It can be evaluated in the shape of any given DFT functional. To stay consistent, here we will plot the potential using the assumption of a BLYP exchange correlation potential.

Aim of the improvement of a DFT functional would be to find that exchange correlation potential that would not be affected by the constrained wavefunction fitting against correlated structure factors, independent of the value of λ. This potential, if consistent for various kinds of datasets, would be the best possible candidate to model the exchange correlation with the same accuracy as costly fully correlated methods, if the First *Hohenberg-Kohn* theorem holds. [100] In this study, when using BLYP, it is quite obvious that this is not the case, which is visualized in Figures 4.6 to 4.8. Since the exchange correlation potential is not confined to have a fixed integral, the calculation of the R$_{RS}$ or PR values would be on an ambiguous scale. For this reason, and since the potential is linked with the correlated density, only the qualitative effects will be discussed in this subsection.



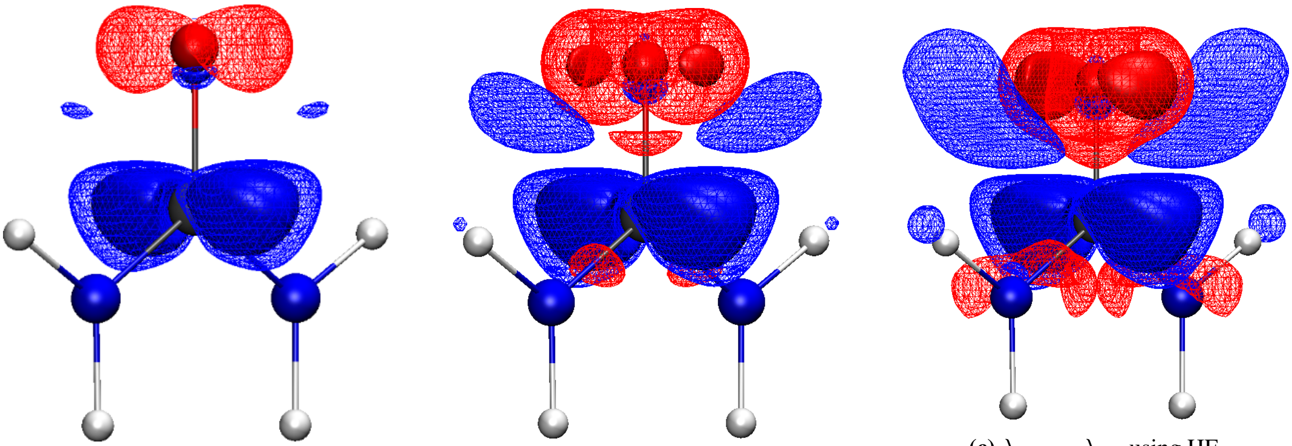
(a) $\lambda_{0.4} - \lambda_{0.0}$ using HF. **(b)** $\lambda_{2.0} - \lambda_{0.0}$ using HF. **(c)** $\lambda_{3.2} - \lambda_{0.0}$ using HF.

(d) $\lambda_{0.4} - \lambda_{0.0}$ using BLYP. **(e)** $\lambda_{1.2} - \lambda_{0.0}$ using BLYP. **(f)** $\lambda_{2.0} - \lambda_{0.0}$ using BLYP.

Figure 4.6: Plot of the difference in electron correlation potential in form of BLYP functional for different λ values during the XCW using structure factors of a resolution below 0.7 Å^{-1}. Individual details are given in subcaptions. Isosurface values in a.u. (solid/wireframe): +0.007/+0.005 (blue) and -0.007/-0.005 (red).

When using a HF-*ansatz* the wavefunction quickly fits the qualitative effect of the exchange correlation potential. The areas only increase in values, the shape of the effect is already observable in the first step of λ when compared to the highest value reached. This coincides with the already quite significant drop in agreement statistics when looking at Figure 4.5 (a), which shows how fast the electron density already agrees with the CCSD density. This is expected, since the exchange correlation DFT is constructed in a way to mimic the exact values, but no exact solution is found, yet.

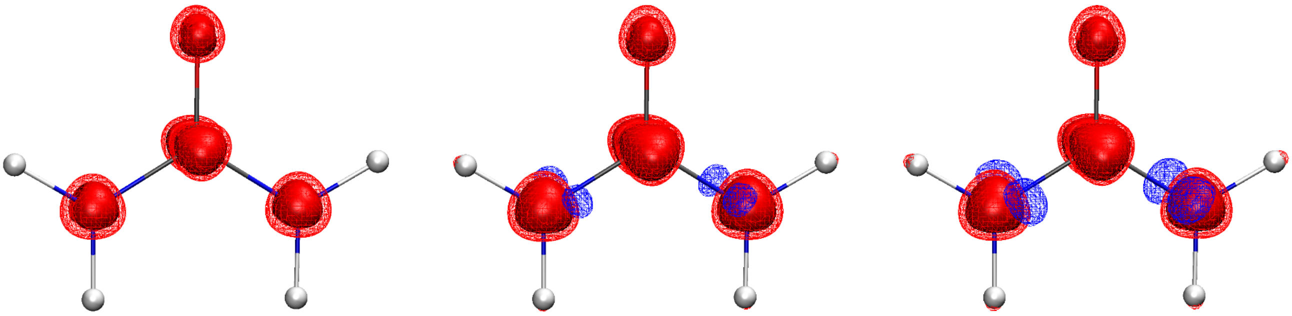
Using an approximated exchange correlation functional, like BLYP, yields an effect of the fitting opposite in sign to what is observed in the HF-*ansatz*. This means the fitting procedure perturbing the wavefunction by the CCSD derived structure factors acts in the opposite direction in the potential as it does for HF. This agrees with the observation that the XCW-fitting brings down over-exaggerated assumptions about the exchange correlation potential of the DFT-functional to the "correct" description as seen in the density agreement and the reconstruction percentage in the previous section. This information can only come through the input into the wavefunction by perturbation through the structure factors that were generated by the use of CCSD.



(a) $\lambda_{2.5} - \lambda_{0.0}$ using HF.

(b) $\lambda_{5.0} - \lambda_{0.0}$ using HF.

(c) $\lambda_{10.0} - \lambda_{0.0}$ using HF.



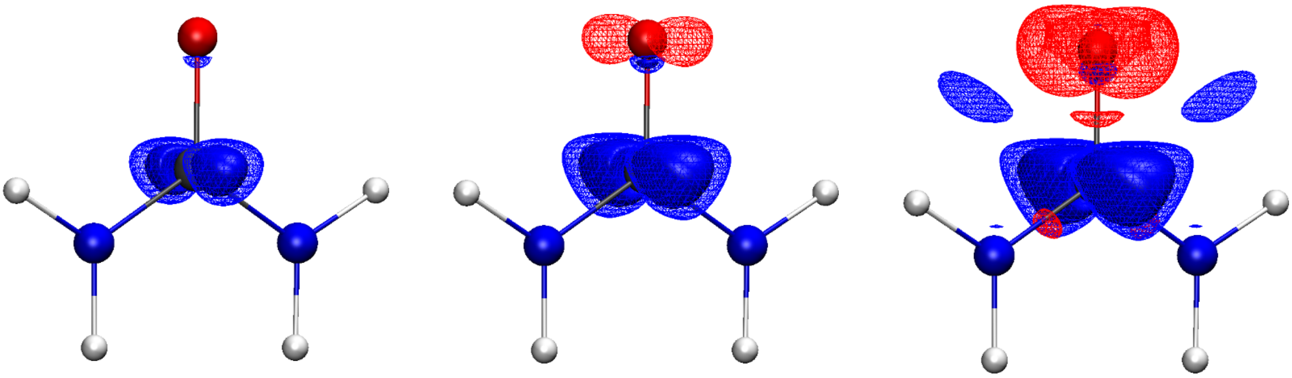
(d) $\lambda_{2.5} - \lambda_{0.0}$ using BLYP.

(e) $\lambda_{5.0} - \lambda_{0.0}$ using BLYP.

(f) $\lambda_{9.5} - \lambda_{0.0}$ using BLYP.

Figure 4.7: Plot of the difference in electron correlation potential in form of BLYP functional for different λ values during the XCW using structure factors of a resolution below 1.4415 Å^{-1}. Individual details are given in subcaptions. Isosurface values in a.u. (solid/wireframe): +0.007/+0.005 (blue) and -0.007/-0.005 (red).

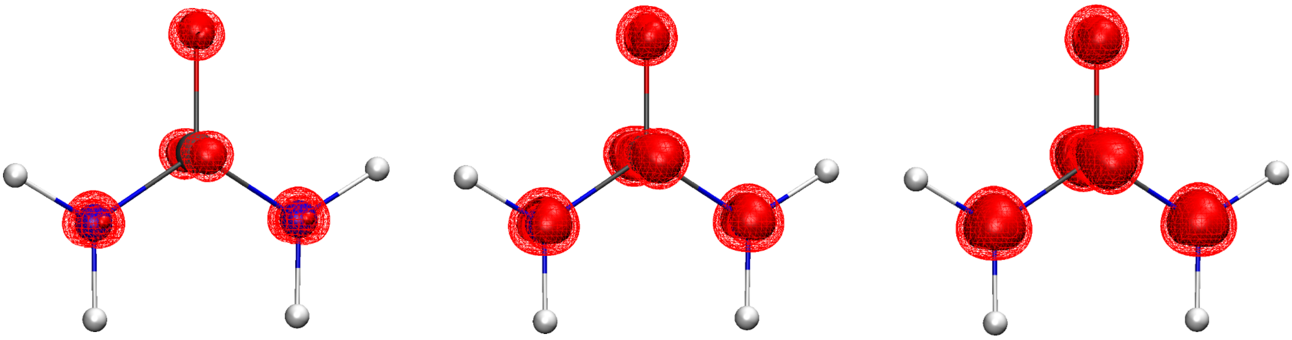
With increasing resolution of included data (see the series of Figures 4.6, 4.7 and 4.8) the same trend as for the ED is observed. The same amount of effect is only fitted at much higher λ values. This shows again that the introduction of higher resolution data did not change the qualitative distribution of the effect. Higher perturbation multiplication is required to reach the same degree of effective perturbation, since in the higher resolution not much new information is introduced, as most reflections already quite well agree with the model (see also section 3.1.2). Identically to the low resolution test case, the wrong assumptions of the BLYP functional are corrected in the DFT-XCW from the structure factors in the higher resolution datasets.



(a) $\lambda_{2.5} - \lambda_{0.0}$ using HF.

(b) $\lambda_{5.0} - \lambda_{0.0}$ using HF.

(c) $\lambda_{10.0} - \lambda_{0.0}$ using HF.



(d) $\lambda_{2.5} - \lambda_{0.0}$ using BLYP.

(e) $\lambda_{5.0} - \lambda_{0.0}$ using BLYP.

(f) $\lambda_{10.0} - \lambda_{0.0}$ using BLYP.

Figure 4.8: Plot of the difference in electron correlation potential in form of BLYP functional for different λ values during the XCW using structure factors of a resolution below 2.0 Å^{-1}. Individual details are given in subcaptions. Isosurface values in a.u. (solid/wireframe): +0.007/+0.005 (blue) and -0.007/-0.005 (red).

4.1.3 Polarization

The polarisation of a molecule due to the crystal field effect can in principle only be correctly included in the wavefunction through periodic boundary calculations (PBC). These require enormously time consuming k-point sampling calculations. In the case of, for example, correlated wavefunction methods, these calcualtion can become far more demanding in terms of computational power than commonly available, maybe even any, existing computer or calculation system could handle. Therefore, the use of high level PBC for the calculation of the wavefunction is tried to be avoided wherever possible. To describe the polarization of the crystal field it has proven successful to use wavefunctions embedded in a model of the surrounding environment, for example by point charges or a combination of charges and dielectric constants present in a crystal to describe the crystal system. [460, 461]
In this work, the reference of the input crystal effect is a periodic calculation of a wavefunction of urea using *CRYSTAL14* with a HF method. HF was chosen to not mix effects of electron correlation into the calculation, which were discussed above. The same calculation was repeated for the molecular wavefunction without PBC and the difference plotted in Figure 4.9 as the reference picture to also compare qualitative distribution in the molecule.

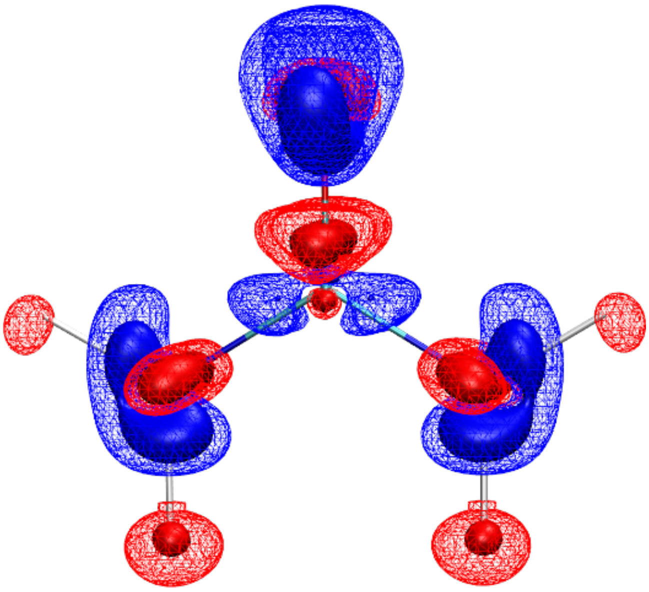
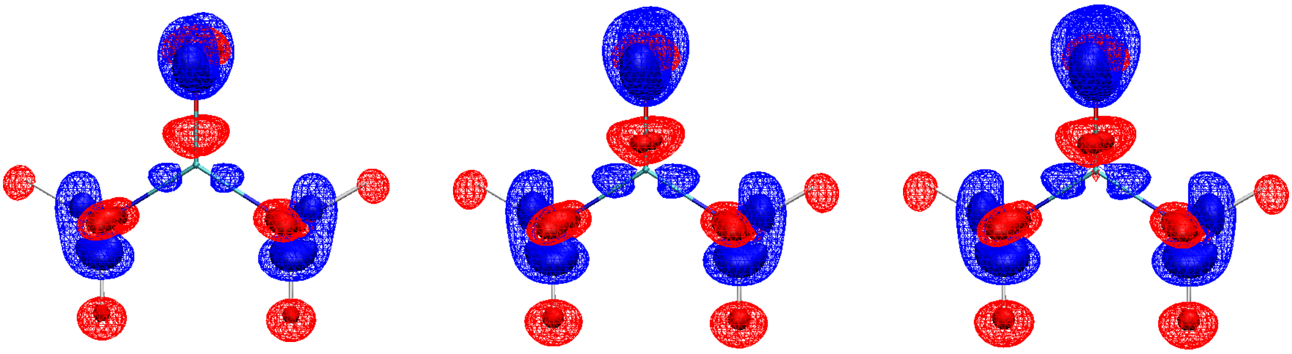


Figure 4.9: Plot of the difference between periodic boundary condition and HF wavefunction density used as reference for the evaluation of reconstruction from XCW fitting. Isosurface values (solid/wireframe): +0.0025/+0.00125 (blue) and -0.0025/-0.00125 (red).

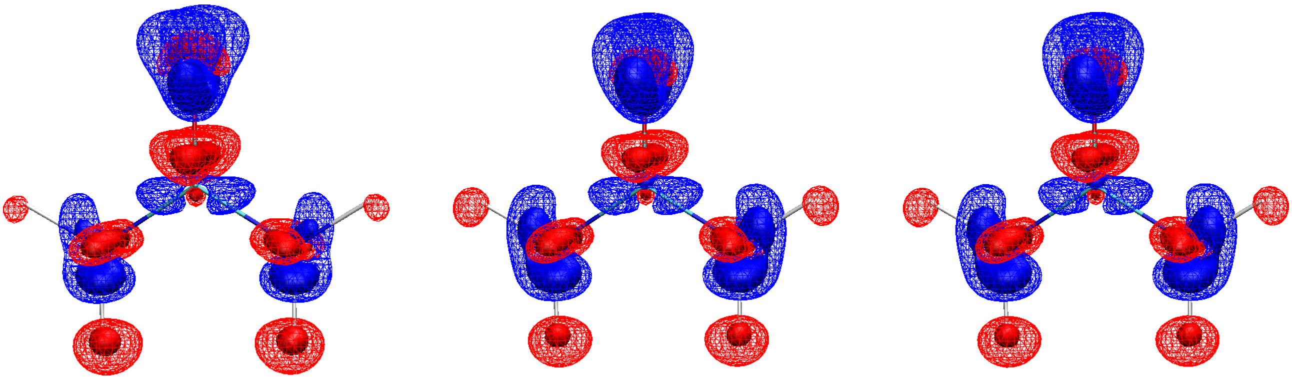
The two XCWs for this case are a pure HF calculation and a calculation using the Hirshfeld charges and dipoles of the calculated HF wavefunction in a self consistent cluster of charges to mimic the crystal field. This model will be referenced as HF-CC. The first fitting at lowest resolution is shown in Figure 4.10.



(a) $\lambda_{0.2} - \lambda_{0.0}$ using HF. PR = 69.9%

(b) $\lambda_{0.8} - \lambda_{0.0}$ using HF. PR = 88.1%.

(c) $\lambda_{1.6} - \lambda_{0.0}$ using HF. PR = 94.0%.



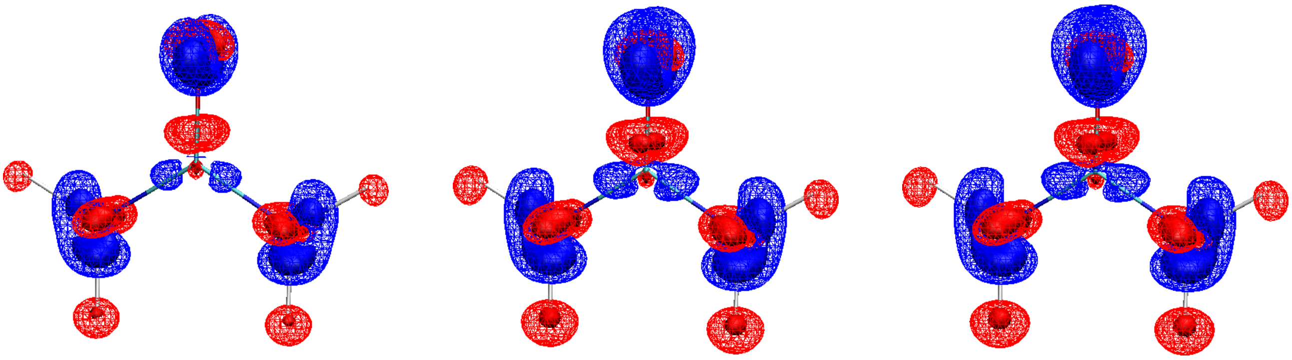
(d) $\lambda_{0.0} - \lambda_{0.0}$ using HF-CC/HF PR = 115.1%.

(e) $\lambda_{0.8} - \lambda_{0.0}$ using HF-CC/HF. PR = 116.7%.

(f) $\lambda_{1.6} - \lambda_{0.0}$ using HF-CC/HF. PR = 116.2%.

Figure 4.10: Plot of the difference in electron correlation potential in form of BLYP functional for different λ values during the XCW using structure factors of a resolution below 0.7 Å^{-1}. Individual details are given in subcaptions. Isosurface values in eÅ^{-3} (solid/wireframe): +0.0025/+0.00125 (blue) and -0.0025/-0.00125 (red).

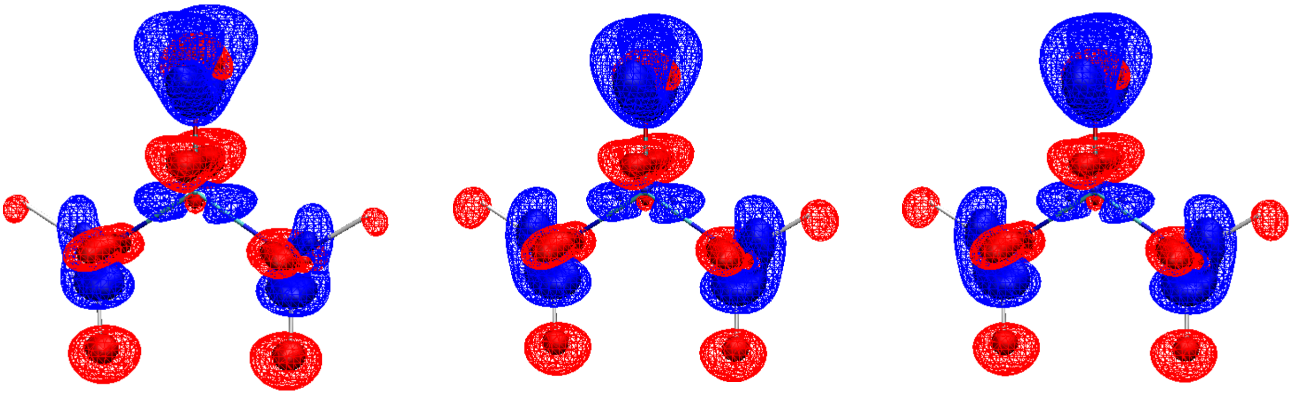
The increasing similarity between the reference desirable effect and the reconstruction through fitting can be seen for the HF fitting, which is reflected also by the increase in PR and decrease in R$_{RS}$. However, the HF-CC fitting looks already complete at $\lambda = 0$. Only minor adjustments of the shape of the difference in ED is obtained by increasing λ. Unfortunately these changes show no improvement compared to the reference calculation and deviate further from the 100% reconstruction value.



(a) $\lambda_{1.0} - \lambda_{0.0}$ using HF. PR = 62.3%.

(b) $\lambda_{5.0} - \lambda_{0.0}$ using HF. PR = 86.0%.

(c) $\lambda_{9.5} - \lambda_{0.0}$ using HF. PR = 92.1%.



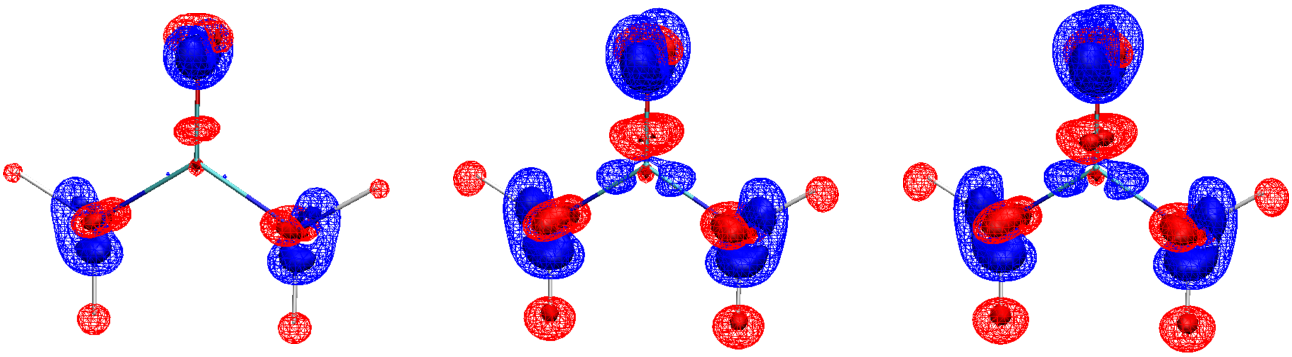
(d) $\lambda_{0.0} - \lambda_{0.0}$ using HF-CC/HF.
PR = 115.1%.

(e) $\lambda_{5.0} - \lambda_{0.0}$ using HF-CC/HF.
PR = 117.8%.

(f) $\lambda_{9.5} - \lambda_{0.0}$ using HF-CC/HF.
PR = 116.2%.

Figure 4.11: Plot of the difference in electron correlation potential in form of BLYP functional for different λ values during the XCW using structure factors of a resolution below 1.4415 Å^{-1}. Individual details are given in subcaptions. Isosurface values in eÅ^{-3} (solid/wireframe): +0.0025/+0.00125 (blue) and -0.0025/-0.00125 (red).

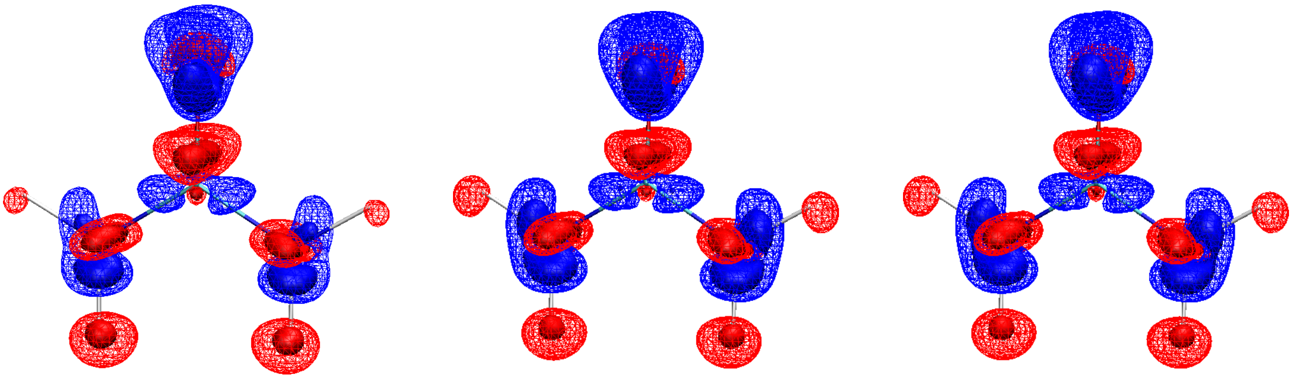
At the experimental resolution it becomes visible that the reconstruction from the structure factors becomes slower with respect to λ, if one watches nitrogen and oxygen during the fitting. The hydrogen atoms seem to be already quite polarized at the first plotted value. Again, the correction of the some differences in shape observed for HF-CC is the only visible difference introduced by fitting the wavefunction.



(a) $\lambda_{1.0} - \lambda_{0.0}$ using HF. PR = 43.7%.

(b) $\lambda_{5.0} - \lambda_{0.0}$ using HF. PR = 73.3%

(c) $\lambda_{10.0} - \lambda_{0.0}$ using HF. PR = 82.8%



(d) $\lambda_{0.0} - \lambda_{0.0}$ using HF-CC/HF. PR = 115.1%

(e) $\lambda_{5.0} - \lambda_{0.0}$ using HF-CC/HF. PR = 117.8%

(f) $\lambda_{10.0} - \lambda_{0.0}$ using HF-CC/HF. PR = 117.1%

Figure 4.12: Plot of the difference in electron correlation potential in form of BLYP functional for different λ values during the XCW using structure factors of a resolution below 2.0 Å^{-1}. Individual details are given in subcaptions. Isosurface values in eÅ^{-3} (solid/wireframe): +0.0025/+0.00125 (blue) and -0.0025/-0.00125 (red).

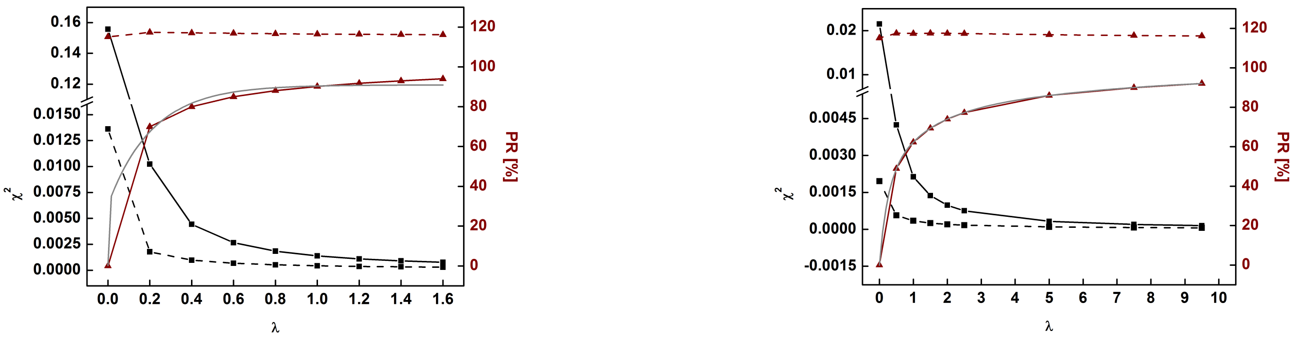
Even at the very high resolution of 2.0 Å^{-1}, where one would expect effects like atom core polarizations to be sufficiently included in the structure factors, if present, no improvement of the HF-CC wavefunction is observed. The HF-XCW can, at comparably high values of λ, recover the majority of the effect. The plot of development of agreement statistics of the fittings is given in Figure 4.13.

The PR, plotted in Figure 4.13, does not significantly change throughout the fitting in the case of using cluster charge wavefunctions. Also, χ^2 does not decrease as drastically as in the HF calculation or also for the modeling of EC (compare Figure 4.5). In this case, the fitting did become unstable at values of $\lambda = 1.6$, which is comparably low with regard to the calculations for EC or also for the higher resolutions (Figures 4.11 and 4.12). Similar to the effect observed for electron correlation, the reconstruction of the polarization from structure factors through XCW becomes slower with regard to λ at this higher resolution. Since the calculated PR for the HF-CC-XCW is not changing in a way which would yield a reasonable trend, the fitting of an exponential function was omitted. The values fitted for the HF-XCW (compare equation 4.1) are given in Table 4.2.

Table 4.2: Fitted coefficients of exponential function for reconstructed interaction density (see equation 4.1).

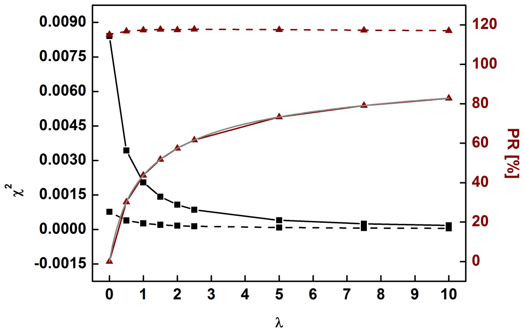
resolution /Å^{-1}	method	a_1	b_1	a_2	b_2	a_3	b_3	c
0.7	HF	-30.13	0.21	-30.13	0.21	-30.49	0.0026	90.97
1.4415	HF	-26.91	5.16	-32.24	0.97	-37.20	0.22	96.35
2.0	HF	-35.81	6.58	-33.99	1.24	-20.85	0.33	90.65

From the exponents of Table 4.2 and the visual trend of Figure 4.13, the fast decay of χ^2 indicates a very quick convergence for the fitting of polarization from the dataset cut to 0.7 Å^{-1}. From the significantly higher exponential coefficients b_i of the higher resolution datasets it is quite clear that these datasets contain less



(a) (sin $\theta/\lambda)_{max}$ = 0.7 Å^{-1}

(b) (sin $\theta/\lambda)_{max}$ = 1.4415 Å^{-1}



(c) (sin $\theta/\lambda)_{max}$ = 2.0 Å^{-1}

Figure 4.13: Graphs of structure factor and electron density agreement throughout XCWs at all resolutions. χ^2 (black) and the PR (red) plotted against the value of λ during the HF-XCW(solid) or HF-CC-XCW(dashed). Additionally exponential functions were fitted to the values to estimate the convergence limit of the PR (gray).

information about polarization per value of λ.

4.1.4 Derivation of the Effects from Experimental Data

If one uses experimental data measured by X-ray diffraction experiments for HF-XCW, the fitted effects are a superposition of EC, polarization and remaining effects. One could formulate the difference in density between λ_0 and λ_{max} in the following way:

$$\Delta\rho = \rho_{\lambda_{max}} - \rho_{\lambda_0} = \Delta\rho_{EC} + \Delta\rho_{pol} + \Delta\rho_{rest}. \tag{4.2}$$

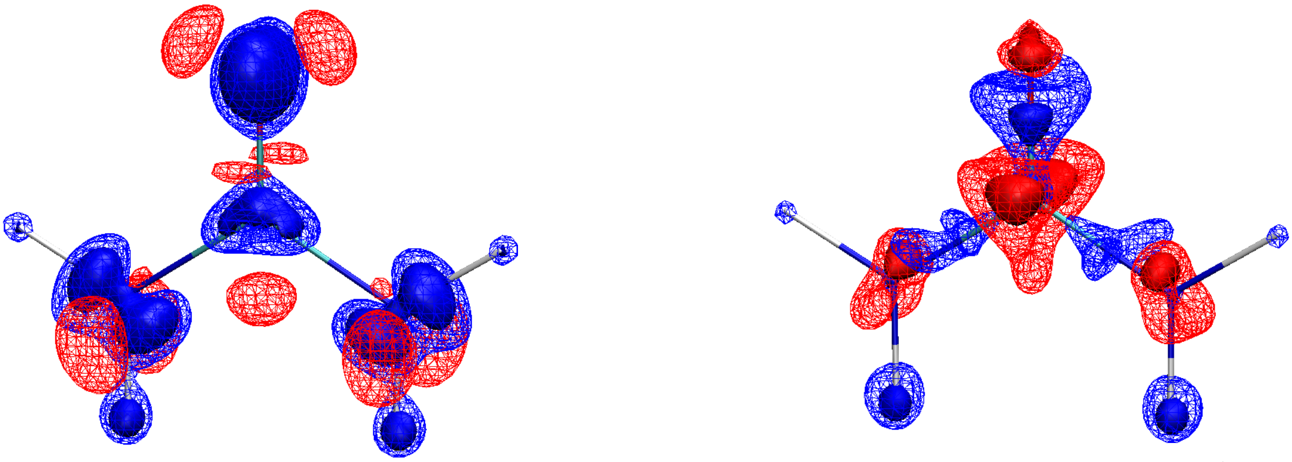
Here, $\Delta\rho_{rest}$ will contain information about untreated experimental errors, not exclusively limited to untreated absorption, extinction, relativistic effects, measurement and machine errors, thermal diffuse scattering, not accurately modeled anharmonic atomic motion etc. Since it was shown above that e.g. a DFT-XCW or HF-CC-XCW can already treat the effect of EC and polarization, respectively, to some degree with $\lambda = 0$, one could reformulate the two contributions of EC and polarization densities. The first part is the "true" difference, with respect to pure HF, and a second part, which is the model over expression or incorrectness:

$$\Delta\rho_{EC} = \Delta\rho_{EC,true} + \Delta\rho_{EC,model} \tag{4.3}$$

and

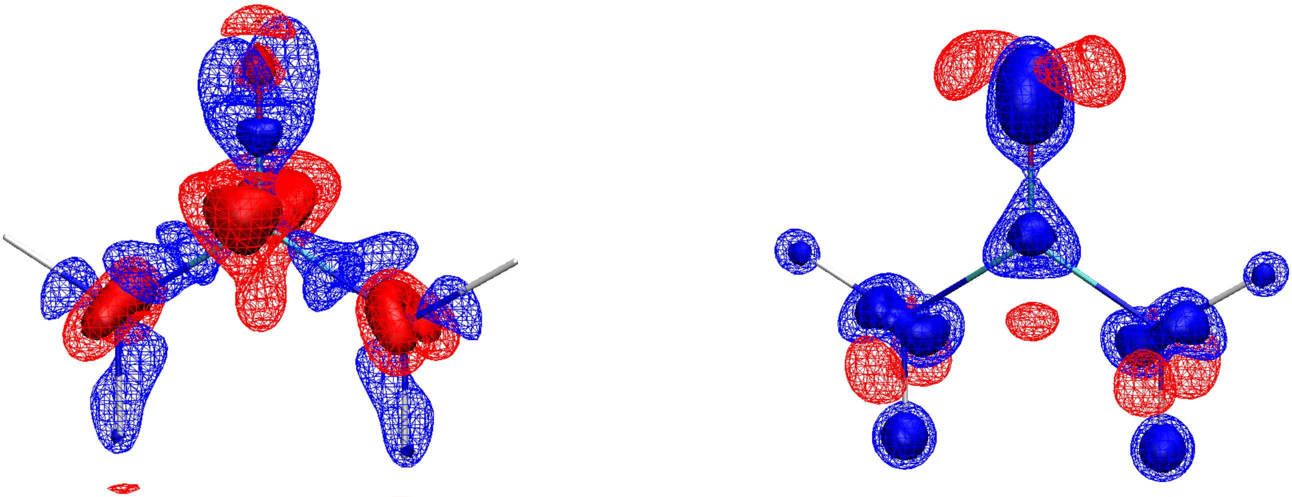
$$\Delta\rho_{pol} = \Delta\rho_{pol,true} + \Delta\rho_{pol,model}. \tag{4.4}$$

Since $\Delta\rho_{EC,true}$ and $\Delta\rho_{pol,true}$ in this model decomposition of density differences are intrinsic properties of the molecule that are indifferent to the description they can be assumed to be constant throughout all procedures. This would correspond to the "true" physical effect. The model components to the difference density, that is in both cases the over-fitting of correlation and polarization, corresponds to the second terms $\Delta\rho_{EC,model}$ and $\Delta\rho_{pol,model}$. From the calculations where wavefunctions were fitted against theoretically derived structure factors we know it is possible to reduce $\Delta\rho_{EC,model}$ to some degree when using XCW. This was quantified by the reduction of χ^2 and increasing PR values using a HF-XCW. In the case of DFT-XCW it is possible to minimize $\Delta\rho_{EC,model}$, but the fitted exponential curve never actually reached 100% reconstruction. Therefore, it is assumed that $\Delta\rho_{EC,model}$ is always present and not 0. For HF-CC-XCW $\Delta\rho_{pol,model}$ shows to be more or less constant and indifferent to subsequent perturbation through XCW fitting. These insights lead to the assumption that the difference electron density of a BLYP-CC-XCW fitting are mainly consisting of two parts: $\Delta\rho_{EC,model}$ and $\Delta\rho_{rest}$.



(a) HF-XCW $\lambda_{max} = 0.0078$; $\Delta\rho = \Delta\rho_{EC,true} + \Delta\rho_{EC,model} + \Delta\rho_{pol,true} + \Delta\rho_{pol,model} + \Delta\rho_{rest}$
All effects.

(b) BLYP-CC-XCW $\lambda_{max} = 0.0049$; $\Delta\rho = \Delta\rho_{EC,model} + \Delta\rho_{rest}$
Experimental and remaining effects.

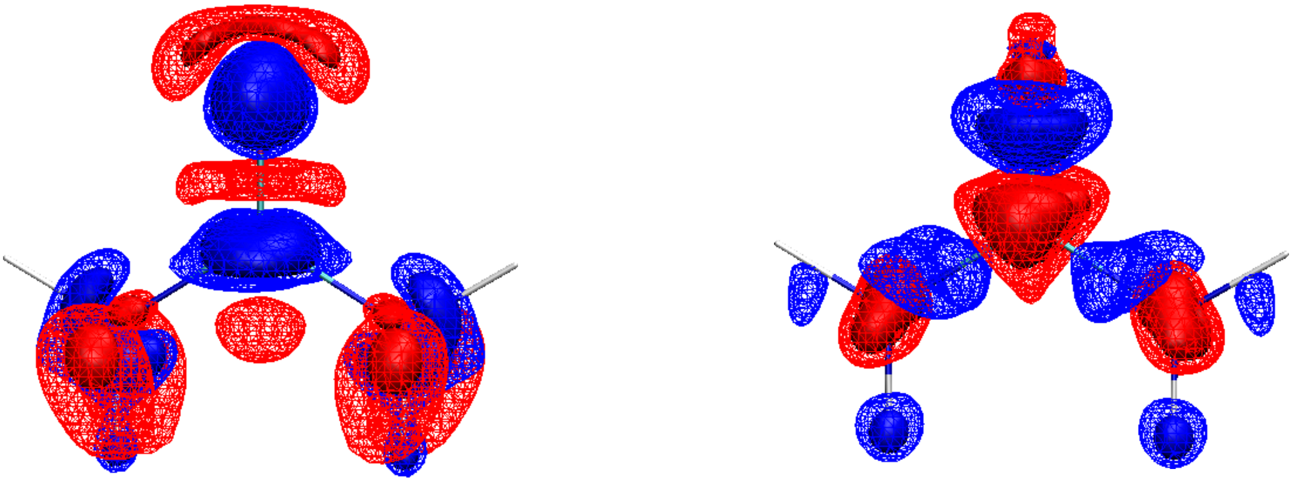


(c) BLYP-XCW $\lambda_{max} = 0.0049$; $\Delta\rho = \Delta\rho_{EC,model} + \Delta\rho_{pol,true} + \Delta\rho_{rest}$
Correlation overfit + polarization + remaining effects.

(d) HF-CC-XCW $\lambda_{max} = 0.0078$; $\Delta\rho = \Delta\rho_{EC,true} + \Delta\rho_{EC,model} + \Delta\rho_{rest}$
Complete correlation + remaining effects.

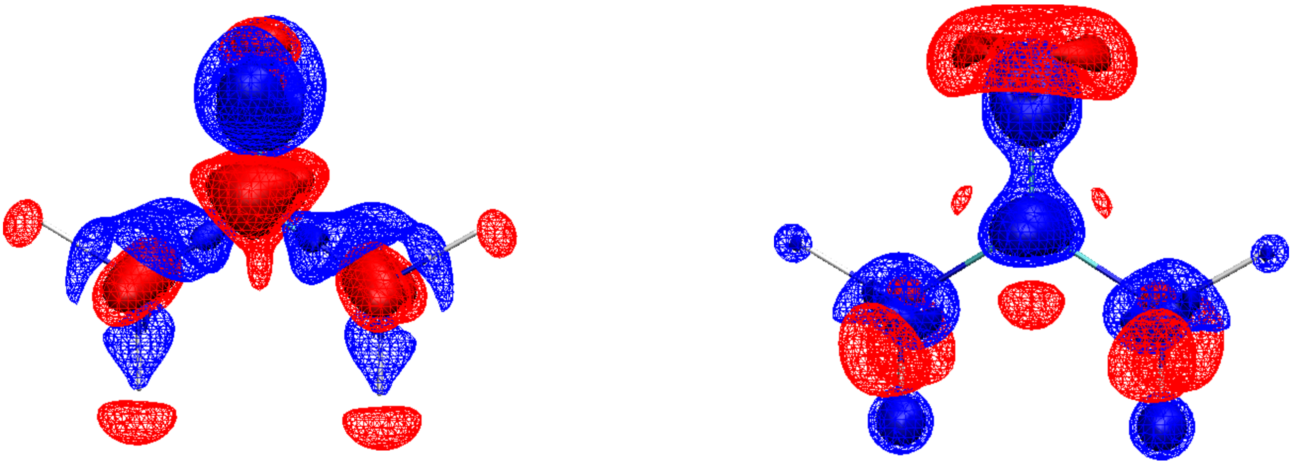
Figure 4.14: Plot of the difference in ρ XCWs using different a *ansatz* based on experimental structure factors of a resolution below 0.7 Å^{-1}. Maximum value of λ reached is shown in subcaptions. Isosurface values in eÅ^{-3} (solid/wireframe): +0.0025/+0.00125 (blue) and -0.0025/-0.00125 (red).

This way of thinking also allows to check whether a combination of methods used in the XCW is able to yield the remaining effect from the difference density in comparison to the theoretical models by qualitative comparison. Therefore HF-XCW was performed, which should include all effects mentioned above as there is no effect included in the method. A BLYP-CC-XCW should in principle yield only the superposition of $\Delta\rho_{EC,model}$ and $\Delta\rho_{rest}$, as all other parameters are modeled in the selected method. BLYP-XCW should leave the polarization and other effects, which should yield the sum of $\Delta\rho_{pol,true}$, $\Delta\rho_{EC,model}$ and $\Delta\rho_{rest}$. $\Delta\rho_{pol,model}$ is assumed to be absent, since there is no model assumption of the polarization in the selected level of theory. As the last model HF-CC-XCW, modeling only the polarization, will give rise to the sum of $\Delta\rho_{EC}$ and $\Delta\rho_{rest}$. The plots of these difference densities at the respective maximal value of λ achievable against the starting point using the pure, unperturbed method at a cut resolution of 0.7 Å^{-1} and the full available resolution is shown in Figures 4.14 and 4.15, respectively.



(a) HF-XCW $\lambda_{max} = 0.046$; $\Delta\rho = \Delta\rho_{EC,true} + \Delta\rho_{EC,model} + \Delta\rho_{pol,true} + \Delta\rho_{pol,model} + \Delta\rho_{rest}$
All effects.

(b) BLYP-CC-XCW $\lambda_{max} = 0.022$; $\Delta\rho = \Delta\rho_{EC,model} + \Delta\rho_{rest}$
Experimental and remaining effects.



(c) BLYP-XCW $\lambda_{max} = 0.022$; $\Delta\rho = \Delta\rho_{EC,model} + \Delta\rho_{pol,true} + \Delta\rho_{rest}$
Correlation overfit + polarization + remaining effects.

(d) HF-CC-XCW $\lambda_{max} = 0.046$; $\Delta\rho = \Delta\rho_{EC,true} + \Delta\rho_{EC,model} + \Delta\rho_{rest}$
Complete correlation + remaining effects.

Figure 4.15: Plot of the difference in ρ XCWs using different *ansatz* selections based on experimental structure factors of a resolution below 1.4415 Å^{-1}. Maximum value of λ reached is shown in subcaptions. Isosurface values in $e\text{Å}^{-3}$ (solid/wireframe): +0.0025/+0.00125 (blue) and -0.0025/-0.00125 (red).

The difference density in Figure 4.15 c) qualitatively resembles the plots of the theoretical polarization models (compare to same resolution in Figure 4.11). Figure 4.15 d) coincides with positive and negative regions of Figure 4.3, which shows that the general model assumption of fitting one effect to obtain the other from the experimental structure factors seems feasible and that experimental errors are not too big that they would bias the fitting too much, comparing Figure 4.15 c) and d) to b).

4.1.5 Discussion of Results

This study confirms the findings by previous theoretical investigations on the electron correlation and polarization fitted by XCW. [455, 457] Also the same kind of resolution dependence of the retrievable effects at a given value of λ is observed. However, in contrast to previous investigations, it is shown that the fitting is not less complete in the case of higher resolution data being used, the value of λ has to be adjusted accordingly in order to obtain the same level of perturbation of the wavefunction calculation. Also an important role for this type of investigation is the choice of uncertainties for the calculated structure factors, since their value directly scales χ^2 and therefore also the required value of λ to introduce the same level of effect by the X-ray constraint.

The application of a quantum mechanical *ansatz*, which describes known effects *a priori* increases the agreement with theoretically calculated structure factors already significantly. Also, it is shown that XCW is able to compensate for an over estimation of the exchange correlation by BLYP when incorporating the structure factors. The maximum agreement obtainable, extrapolated by the exponential fits in the value of c of the PR curves according to equation 4.1 for the BLYP-XCW shows no systematic trend. On the other hand, the exponents show a significant increase with resolution. This reflects the increased values of λ required to obtain the same amount of reconstruction from structure factors.

Possible improvement of XCW

It would be desirable to eliminate the resolution dependence of λ in the XCW fitting and to obtain a similar level of perturbation at any resolution of the dataset used. One way to achieve this might be the selection of an intensity or significance difference threshold for the selection of a subset of reflections to be used for the fitting. The subset would be selected by for example the significance of differences between measured and observed intensities in terms of the uncertainty of the measured data. This would only introduce a portion of reflections for the calculation of χ^2 to use during XCW fitting and therefore easily provide a set of reflections to check for over fitting by XCW, if χ^2 of the remaining reflections is monitored. If the agreement to all other reflections becomes systematically and significantly worse, the fitting could be considered to introduce noise or effects not represented in the whole dataset.

The proposed decomposition of the difference density into contributions of EC, polarization and other effects allows for the experimental data sets to be analysed through exclusion of the individual effects and to see all effects that would be expected in terms of difference densities are also retrievable from the experimental data. One could try to obtain individual contributions of all effects by different linear combinations of the calculated densities. However, this was beyond the scope of this work at this point.

Implications for HAR

If DFT is used instead of HF, the majority of electron correlation effects retrievable by XCW is already described. Therefore a HAR using DFT, in an ideal case a more sophisticated or precise exchange correlation functional than BLYP, should yield the best structural description, as the wavefunction will describe the molecular density better. Although the effect is overexpressed (156% of the effect expected compared to CCSD), this is only 56% away from the correct effect, while the HF wavefunction misses 100% of the effect.

The cluster charges implemented in *Tonto* can provide a very sophisticated description of the environment in the crystal, but only if the interaction between the wavefunction and its environment is mainly electrostatic. Therefore using cluster charges in HAR should in principle improve the structural description, as the polarization of the wavefunction is much better, even if overestimated (20 % over-estimation vs. 100% under-estimation without cluster charges). If there is covalent contribution – for example a very strong hydrogen bond or other strong interactions which are not driven by pure electrostatic polarization of the wavefunction – the cluster charges will not be able to correctly describe the situation. Here, the best approach would require explicit neighbours in the wavefunction calculation or periodic boundary conditions.

Possible extension using electron diffraction

Since it was shown recently that the transformation of non-spherical X-ray form factors into electron diffraction form factors is yielding reasonable and more accurate results compared to the IAM, [462] it might be of interest to investigate whether the implementation of XWR in the context of electron diffraction might yield even more insight into effects like the exchange correlation potential. This is because electron diffraction is based on the diffraction from electrostatic potentials, rather than electron density itself, therefore the exchange correlation potential might become a fittable parameter from electron diffraction using the methodology presented above for electron diffraction. However, this relies on a reasonably good description of the experimental errors and improvement of precision of experimental results.

4.2 *Publication:* Chemical Bonding in Polarised Push-Pull Ethylenes

My contribution to this paper was the measurement of the three crystal structures, for which I subsequently performed XWR. The crystals were prepared by the group of Prof. Hikaru Yanai in Tokyo, Japan and sent to me for analysis. For the first time ever, this type of analysis was performed on a twinned crystal. The paragraph starting with "Subsequently, HAR and X-ray-constrained..." up to the concluding paragraph were written by me. Tables 2, 3 and Figure 5 were generated by me. All analysis, including NBO, ELI, QTAIM, RGBI and EDA, were done by me. The idea for deconvolution of the EDA by using model compounds was also proposed and performed by me. The supporting information containing more data and tables about the bonding analysis, as well as a description of the EDA can be found in Appendix C.1. The Complementary Bonding Analysis was performed based on the *.ffn* files introduced during this work and subsequent conversion and calculations of properties using *cuQCrT* (see section 3.3).

This publication shows that performing XWR on in-house data can provide already a full insight into the chemical bonding situation in a molecule if the crystals scatter with sufficient intensity and quality. The fact that even a twinned crystal can be used is promising for a broader applicability of the method, given that the de-twinning is sufficiently well performable. Especially in the context of the reconstruction of correlation and polarization from structure factors using XWR and their dependence on the resolution (see section 4.1), this type of investigation is expected to yield sufficient results from this type of diffraction experiment.

The presented procedure allows to perform a broad spectrum of X-ray constrained bonding analysis on in-house data that provides results of high quality. Pure DFT failed to predict the correct bonding situation and geometry in *ab initio* calculations and only perturbation theory was able to describe the bonding situation of these compounds as correctly as XWR. This case is a proof that XWR can in principle be applied in a broad field to enhance insights into chemical bonding from in-house diffraction data, that would otherwise only be accessible through very time consuming and sophisticated theoretical calculations.

Reprinted with permission from *Angewandte Chemie International Edition* **2019**, *58*, 26, 8839-8844 (reference number: 4897731001454).

GDCh Communications Angewandte International Edition

International Edition: DOI: 10.1002/anie.201904176
German Edition: DOI: 10.1002/ange.201904176

Bonding Analysis

Chemical Bonding in Polarised Push–Pull Ethylenes

Hikaru Yanai, Takumi Suzuki, Florian Kleemiss, Haruhiko Fukaya, Yasuo Dobashi, Lorraine A. Malaspina, Simon Grabowsky,* and Takashi Matsumoto**

Abstract: *1,1-Diamino-2,2-bis(triflyl)ethylenes with both twisted and planar structures around the partial “C=C” bond were synthesised. Bonding properties in these compounds were analysed by an experimental approach using high-resolution X-ray diffraction data treated with X-ray wavefunction refinement (XWR). In the twisted compound, a dominant contribution of the charge-separated resonance structure was revealed. On the contrary, the nearly planar compound still showed π-bonding character, however, with a considerable contribution of the charge-separated resonance structure.*

The C=C double bond is ubiquitously found in organic molecules. According to textbook knowledge, the 2sp^2–2sp^2 head-on overlap forms a σ-bond and the sideways overlap of the 2p-orbitals represents a π-bond.[1] Geometric requirements of π-bonding result in a planar structure for ethylenes, with restricted internal rotation around the C–C axis.[2,3] Push–pull ethylenes, where one carbon atom is substituted by an electron-donating group and the other one by an electron-withdrawing group, are arranged in special ways. In the simple picture, they may be classified into planar (type I) and twisted (type II) compounds (Figure 1). Both examples have been reported with a 2-(2-imidazolidinylidene)-1,3-diketone framework: *N*,*N*′-dihydro derivative **1a**[4a] is a type-I compound with a twisting angle τ of 4.8°;[5] *N*,*N*-diisopropyl derivative **1b** ($\tau = 82.8°$) is regarded as a type-II compound.[4b] An experimental electron-density study of **1b** using Stewart's multipole model confirms that the charge-separated resonance structure **D** is dominant for such twisted compounds.[4b]

The triflyl (Tf = CF$_3$SO$_2$) group is one of the strongest electron-withdrawing groups. As shown by the outstandingly high electrophilicity of Tf$_2$C=CH$_2$,[6] the attachment of triflyl groups to the ethylenic carbon leads to a large polarisation of σ- and π-electrons. Additionally, inter-[7] and intramolecular salts[8] containing the carbanion [Tf$_2$CR]$^-$, which is a conjugate base of the superacidic carbon acid Tf$_2$CHR,[9,10] have been isolated.

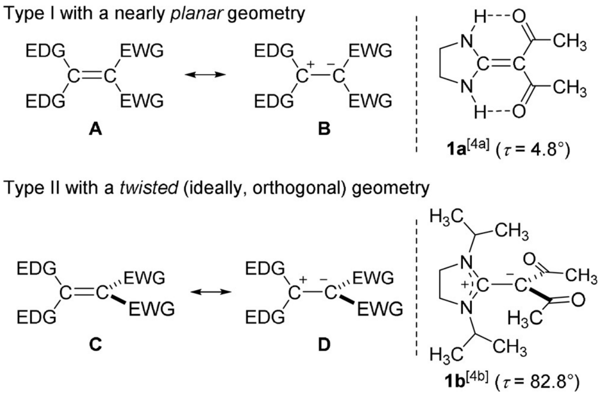


Figure 1. Type-I and type-II push–pull ethylenes (EDG = electron-donating group; EWG = electron-withdrawing group). Compound **1b** is drawn as an acylcarbanion, although it contributes less than the two enolate resonance structures that are likewise possible here, but do not occur with triflyl substitution (see the main text for details).

Quantum-chemical calculations of [Tf$_2$CR]$^-$ have revealed a pronounced carbanionic character, which is stabilised by the delocalisation of a non-bonding electron pair on the anionic carbon to the adjacent $\sigma_{S-C(F_3)}$* orbitals, called negative hyperconjugation.[8b,c] This is different to usual carbanion equivalents such as enolates in which the enolate resonance structure contributes more than the acylcarbanion one. Based on this background, we gained interest in the “C=C” bonds in 1,1-diamino-2,2-bis(triflyl)ethylenes, which can be considered as resonance hybrids between π-bonding Tf$_2$C=C(NHR)$_2$ and charge-separated [Tf$_2$C]$^-$–[C(NHR)$_2$]$^+$. Such ethylenes would present ultimately polarised “C=C” bonds. Although a diisopropyl derivative (R = *i*-Pr) with type-II structure in the solid state was reported,[11] type-I derivatives are unknown. Here we describe the “C=C” bonding properties in both types of 1,1-diamino-2,2-bis(triflyl)ethylenes, revealed by a combination of theoretical and experimental techniques. Polarised push–pull systems are promising motifs for organocatalysts,[12] ligands for metal complexes,[13] and non-linear optical materials.[14] The present work demonstrates a quantitative and reliable understanding of the ultimately polarised olefinic systems, providing basic insight into such chemistry.

According to Hanack's procedure,[11] not only dialkyl derivatives **3** but also diaryl derivatives **4a–d** are obtained by the reactions of Tf$_2$CH$_2$ **2** with carbodiimides (Table 1). It turned out that the substituents completely regulate the conformational preference, so that both type-I (**3a,b**) and type-II (**4a–d**) compounds could be designed. This was

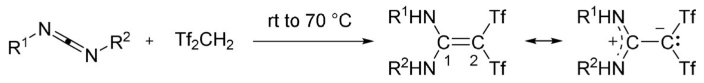
---

[*] Dr. H. Yanai, T. Suzuki, H. Fukaya, Dr. Y. Dobashi,
Prof. Dr. T. Matsumoto
School of Pharmacy, Tokyo University of Pharmacy and Life Sciences
1432-1 Horinouchi, Hachioji, Tokyo 192-0392 (Japan)
E-mail: yanai@toyaku.ac.jp
tmatsumo@toyaku.ac.jp

F. Kleemiss, L. A. Malaspina, Prof. Dr. S. Grabowsky
Institut für Anorganische Chemie und Kristallographie, Fachbereich 2-Biologie/Chemie, Universität Bremen
Leobener Str. 3, 28359 Bremen (Germany)
E-mail: simon.grabowsky@uni-bremen.de

Supporting information and the ORCID identification number(s) for the author(s) of this article can be found under: https://doi.org/10.1002/anie.201904176.

Table 1: Synthesis of **3** and **4** and selected NMR data.

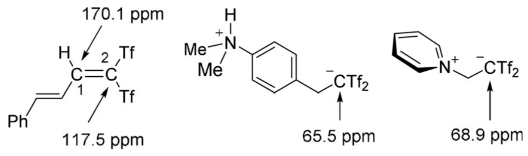


(1.1 equiv) **2** **3**, **4**

Entry	R^1	R^2	**3**, **4**	Yield[a] [%]	δ[b] [ppm] NH	C1	C2	$\Delta\delta_{C1\text{-}C2}$[b] [ppm]
1[c]	*i*-Pr	*i*-Pr	**3a**	66	7.47, 7.02	154.8	67.0	87.8
2[c]	*c*-Hex	*c*-Hex	**3b**	24	7.41, 7.02	154.9	67.3	87.6
3[c]	4-MeC_6H_4	4-MeC_6H_4	**4a**	72	9.33	153.5	74.0	79.5
4[d]	Ph	Ph	**4b**	74	9.45	154.0	74.3	79.7
5[d]	4-$MeOC_6H_4$	Ph	**4c**	91	9.54[e]	153.8	74.0	79.8
6[d]	4-BrC_6H_4	Ph	**4d**	70	9.55, 9.45	154.5	74.6	79.9

[a] Isolated yield. [b] At 25 °C. For ^1H, 0.010 mol L^{-1} solutions; for ^{13}C, 0.20 mol L^{-1} solutions. [c] Reaction was carried out at room temperature in CH_2Cl_2. [d] Reaction was carried out at 50 °C in $ClCH_2CH_2Cl$. [e] Two N–H signals were overlapped.

confirmed by the NMR analysis, which revealed revealed significant differences in the conformation, depending on the substituents. In both **3a** and **3b**, N–H moieties as well as alkyl groups are magnetically inequivalent. On the contrary, ^1H NMR spectra of **4a** and **4b** show conformational symmetry and intramolecular hydrogen bonding between the N–H part and the sulfonic oxygen, which is supported by the fact that no change of the signals is observed in a range from 0.010 to 0.20 mol L^{-1}. Note that the substituents on the nitrogen atoms strongly affect the chemical shift δ_{C2}: Compared with **3a** and **3b**, δ_{C2} in **4a–d** is shifted about 7 ppm downfield. Because the twisted conformer of **3a** bears a s-*cis*,s-*trans* configuration around two C1–N bonds, as found in the crystal structure (see below), the NMR data implies that this conformation is preferred even in CD_3CN solutions. On the contrary, NMR data of **4** demonstrate the preference of a nearly planar s-*cis*,s-*cis* conformer. Comparison of ^{13}C NMR data of *gem*-bis(triflyl)diene **5**[8c] and isolable zwitterions **6**[9b] and **7**[9a] reveals more details about the electronic state (Figure 2). Because δ_{C2} in **3** is close to the values of the anionic carbon atoms in **6** and **7** (65.5 and 68.9 ppm, respectively), the charge-separated resonance structure **D** can be considered the main contributor. Although the downfield shift of δ_{C2} in **4** can be attributed to the increasing double-bond character, the magnitude is still close to the anionic carbon atoms in **6** and **7**. In solid-state IR spectroscopy, the C1–C2 stretching vibration ν_{C1-C2} of **3a** was observed at either 1377 or 1359 cm^{-1} (an unambiguous assignment was not possible). Although the $\nu_{C1\text{-}C2}$ band of **4a** (1374 cm^{-1}) is essentially similar to that of **3a**, it is shifted to a much higher wavenumber (1515 cm^{-1}) in **5**.



5 **6** **7**

Figure 2. ^{13}C NMR chemical shifts of reference compounds **5–7** in CD_3CN.

To obtain detailed structural information, we conducted X-ray diffraction studies of all products and reference

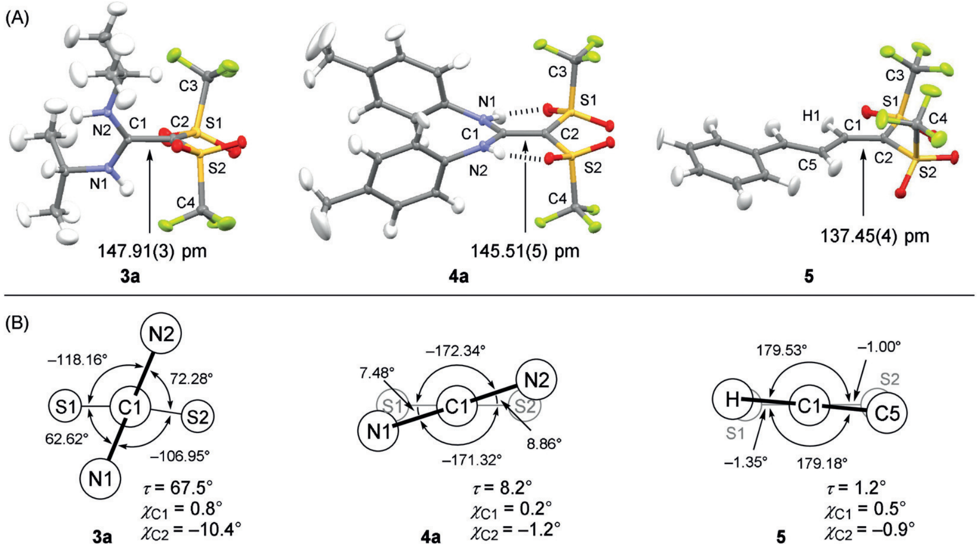


Figure 3. A) HAR-X-ray crystallographic structures of **3a**, **4a**, and **5**; anisotropic displacement parameters at a probability level of 50%. B) Newman projections including τ, χ_{C1}, and χ_{C2}.

compound **5**. For **3a**, **4a**, and **5**, accurate molecular geometries and anisotropic displacement parameters of all atoms including hydrogen atoms were obtained by Hirshfeld atom refinement (HAR) (Figure 3 A),[15,16] which agree well with Hanack's findings:[11] the twisting angle τ is 67.5° in **3a** and the pyramidal parameters χ_{C1} and χ_{C2} are 0.8° and −10.4°, respectively (Figure 3 B).[5] The stereochemistry around the C1−N bonds is s-*cis*,s-*trans.* In contrast, **4a** shows the desired nearly planar, s-*cis*,s-*cis* geometry in the crystal ($\tau = 8.2°$, $\chi_{C1} = 0.2°$, $\chi_{C2} = -1.2°$). The geometry also supports the two-point intramolecular hydrogen bonding between the N−H part and the sulfonic oxygen atoms (Figure 3 A). The C1−C2 distance in **4a** is slightly shorter than the value in **3a** (145.51(5) vs. 147.91(3) pm, respectively). However, the difference is small and both interatomic distances are comparable to the typical length of a $C_{sp^2}-C_{sp^2}$ single bond in buta-1,3-diene (≈ 148 pm).[17] Even in the DFT-optimised structures (M06-2X/6–311 + G(d,p)) of the isolated state, no significant changes of the C1−C2 distances were observed in both cases (see the Supporting Information). DFT calculations mapping the internal rotation around the C1−C2 axis of the diphenyl derivative **4b** reveal a very small rotational barrier ΔH_{rot} (27.2 kJ mol^{-1}; see the Supporting Information).

Subsequently, HAR and X-ray-constrained wavefunction fitting were performed. Here, the experimental information provided by the structure factors is used to construct the wavefunction[18] and it is a tool to obtain the total molecular wavefunction in an experimental manner from the diffraction experiment. The full procedure is termed X-ray wavefunction refinement (XWR).[19] XWR was carried out for **3a**, **4a**, and **5**, that is, experimentally refined wavefunctions were obtained and a quantum crystallographic bonding analysis could be performed.[20] The bond descriptors that were used are the Roby–Gould Bond Index (RGBI),[21] Bader's Quantum Theory of Atoms in Molecules (QTAIM),[22] Kohout's Electron Localizability Indicator (ELI),[23] the Raub–Jansen Index (RJI),[24] Natural Bond Orbital (NBO) analysis,[25] and Natural Resonance Theory (NRT)[26] analysis. Resonance structures that reflect the molecular geometries found in the crystal structures are shown in Figure 4.

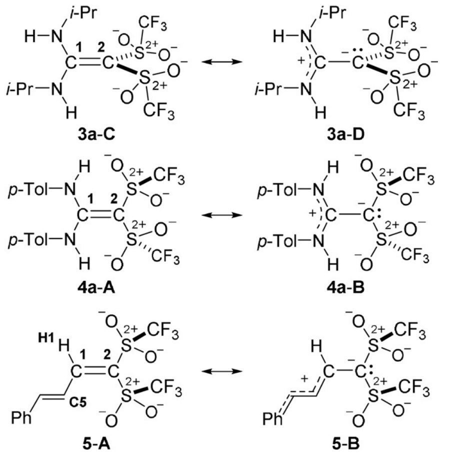


Figure 4. Resonance structures of **3a**, **4a**, and **5** including their geometries in the crystal structures. The resonance structure of the formal "C=C" bond is investigated in this paper, whereas the charge-separated resonance structure of the sulfonyl group was already validated in ref. [19b] based on XWR. In the Supporting Information, we show that also in this case, the sulfur atom is clearly not hypervalent.

Table 2: XWR-derived bonding parameters of **3a**, **4a** and **5**.

	3a	**4a**	**5**
RGBI	1.179	1.335	1.760
Covalent RGBI	1.152	1.315	1.756
Ionic RGBI	0.255	0.233	0.115
Covalency [%]	95.3	97.0	99.5
ρ_{BCP} [e Å^{-3}]	1.831	1.864	2.260
$\nabla^2\rho_{BCP}$ [e Å^{-5}]	−18.48	−17.02	−26.95
ε_{BCP}	0.112	0.232	0.226
DI	0.998	1.139	1.540
N_{ELI}(C1-C2) (*e*)	2.022	2.417	3.553
N_{ELI}(LP$_{C2}$) (*e*)	1.880	1.011	–
RJI ρ_{C1} [%]	60.5	51.3	34.8

All bonding indices in Table 2 show that the C1−C2 bond in **3a** is more similar to a highly polarised single bond (resonance structure **D**), whereas the covalent and double-bond character increase to **4a** and **5**: In terms of the RGBI, the formal "C=C" bonds in both **3a** (1.179) and **4a** (1.335) cannot be considered normal double bonds as in ethylene (2.000) or an electron-deficient double bond as found in **5** (1.760). Compared with **5**, a large ionic RGBI coinciding with a smaller covalent RGBI of **4a** as well as **3a** indicate a highly polarised character of the C1−C2 bonds.

The electron density ρ at the C1−C2 bond-critical point (BCP) obtained from the XWR-based QTAIM analysis increases, and the Laplacian $\nabla^2\rho$ becomes more negative (more covalent character) from **3a** towards **5**. In the Supporting Information, a plot of the profile of $\nabla^2\rho$ along the C1−C2 bond axis is shown that supports this trend. Considering the ellipticity ε of an isolated C−C single bond (0 in ethane), a 1.5-fold C−C bond (0.23 in benzene), and an isolated C=C double bond (0.45 in ethylene), the ε value of 0.112 in **3a** indicates that the C1−C2 bond in this system can be placed between a single and a 1.5-fold bond. The ε value of **4a** (0.232) is close to the value of the 1.5-fold bond. The delocalisation index (DI),[27,28] which may be regarded as a covalent bond order, is close to the value of a pure single bond (1.0) in **3a** and increases from **4a** to **5**. In the context of the QTAIM analysis, it is interesting that the two intramolecular N−H⋯O hydrogen bonds in compound **4a** (Figure 3 A) are strong and include covalent contributions, indicated by significant electron-density accumulations at the H⋯O bond-critical points ($\rho_{BCP} = 0.258$ and 0.263 e Å^{-3} with $\nabla^2\rho_{BCP} = 2.78$ and 2.53 e Å^{-5}, respectively). These are signs of resonance-assisted hydrogen bonds which close a six-membered ring here and facilitate electron delocalisation in the system.[29]

We have complemented these findings with an energy decomposition analysis (EDA) based on purely theoretical wavefunctions of the interacting fragments (see the Supporting Information).[30] By a suitable selection and combination

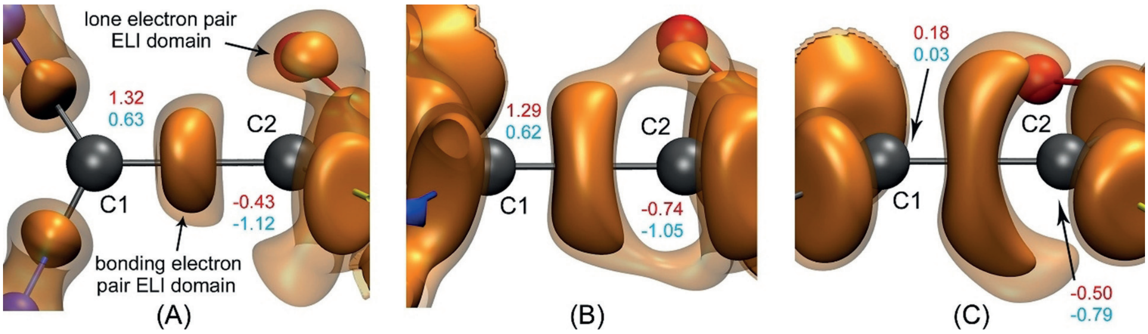


Figure 5. Isosurfaces of the XWR-derived ELI for A) **3a**, B) **4a**, and C) **5**. Isovalue = 1.45 (transparent), 1.60/1.52 (solid, **3a**/[**4a** and **5**]). XWR-derived atomic charges are given in *e* (red, QTAIM charge; blue, NPA charge).

Table 3: XWR-derived NBO parameters of **3a**, **4a** and **5**.

		3a	**4a**	**5**
NLMO/NPA bond order		1.015	1.173	1.525
NBO bond order	*Total*	1.029	1.776	1.931
	Covalent	0.987	1.238	1.567
	Ionic	0.042	0.538	0.364
Resonance weight[a] [%]		3:97[b]	79:21[c]	93:7[c]

[a] Sums of the contributions of the resonance structures. [b] **C**/**D** ratio. [c] **A**/**B** ratio.

of model compounds, we show that the formal "C=C" bond formation exceeds all other interactions between the two fragments by 312 kJ mol^{-1} in **3a**, by 598 kJ mol^{-1} in **4a**—despite the hydrogen-bonded interaction described above that is not present in **3a**—and by 763 kJ mol mol^{-1} in **5**. This is a measure of the importance of the double-bond character that increases from **3a** to **5**, because these energetic trends are caused by the orbital-overlap term, whereas the electrostatic term is approximately constant.

The XWR-derived ELI localisation-domain representations visualise the partial "C=C" double bonds. In the plots of **3a** and **4a** (Figure 5 A,B), localisation domains representing lone-electron-pair ELI basins populated with 1.880 *e* and 1.011 *e*, respectively, are shown above atom C2. In contrast, such a basin is not present in the case of **5** (Figure 5 C). This implies that the charge-separated resonance structures **B** and **D** play a substantial role for **3a** and **4a**, being more pronounced in compound **3a**. The N_{ELI} values of the C1–C2 bond basins in **3a** (2.022 *e*) and **4a** (2.417 *e*) are much lower than that of **5** (3.553 *e*; Table 2). However, the different shape of the domain in **3a** (Figure 5 A, round shape) compared to **4a** (Figure 5 B, elongated shape) hints towards a significantly higher π-bonding contribution in **4a** compared to **3a**, which rather resembles a highly polarised single bond. The RJI of **3a** shows a contribution of more than 50 % of C1 to the C1–C2 bond ELI basin owing to electron depletion at C2 due to triflyl substitution. Compared with **3a**, the bonding situation in **4a** can be understood as a stronger delocalisation of a lone electron pair on C2 (LP$_{\mathrm{C2}}$) towards the bond, although C1 still contributes more than 50 % of the electron density into the ELI basin. Despite relatively weak effects of C5 and H1 to the "C=C" bond (see Figure 3 for atomic labels), diene **5** fits in this series as an example that shows a strong double-bond character, where C1 contributes less than 40 %.

The NBO/NRT analysis quantitatively evaluates resonance contributions linked to the Lewis structure.[25,26] Therefore, we also analysed the XWR wavefunctions with the NBO 6.0 program (Table 3). The NBO/NRT analysis of **3a** reveals that it should unambiguously be considered as a well-defined "carbenium ylide" **3a-D** rather than an "alkene" **3a-C** on the basis of the following results: 1) Natural population analysis (NPA) charges of C1 (0.63 *e*) and C2 (−1.12 *e*; Figure 5); 2) NLMO/NPA C1–C2 bond order of 1.015, close to 1; 3) NBO C1–C2 bond order of 1.029 with a strong covalent character (covalent/ionic = 0.987:0.042); 4) dominance of the charge-separated resonance structure **3a-D** (**C**/**D** = 3:97) in the NRT analysis, and 5) a LP$_{\mathrm{C2}}$ orbital with an occupancy of 1.707 *e*. In contrast, the C1–C2 bond in **5** exhibits an obvious double-bond character (NBO bond order = 1.931, covalent/ionic = 1.567:0.364) and a dominant contribution of the π-bonding resonance structure **5-A** (**A**/**B** = 93:7). A similar bonding analysis of the bond in **4a** reveals unusual and intermediate properties. The NBO program interprets the π-bonding resonance structure **4a-A** still to be the major contributor (**A**/**B** = 79:21), but the resonance weight of the charge-separated one **4a-B** is more than 10 % higher than in **5** and therefore more significant. The result of the XWR wavefunction is in agreement with the result from an MP2 calculation (**A**/**B** = 82:18), whereas a corresponding Hartree-Fock calculation fails to reproduce it (see the Supporting Information). This confirms that electron correlation—here obtained through either perturbation theory or the experimental fitting procedure—is crucial for an accurate bonding analysis.

In conclusion, we found that the XWR wavefunctions successfully describe three different formal "C=C" bonds in *gem*-bis(triflyl)alkenes in the solid state. In **3a**, with a large distortion around the central C–C axis, the data obtained in all bond descriptions as well as ^{13}C NMR and IR spectroscopy support that it exhibits the pronounced character of a lone electron pair on the C2 atom. In other words, **3a** must be represented as a charge-separated carbenium ylide. The nearly 100 % contribution of the charge-separated resonance structure **3a-D** established in the NRT analysis is surprising, because its twisting angle τ (67.5°) is still far from an orthogonal angle (90°). Although the "C=C" bond in diene **5** exhibits some electron deficient properties, it can be categorised as a normal double bond. As an intermediate entry, **4a**, with a nearly planar geometry, shows the character of a partial "C=C" double bond. In most bond indices including RGBI, DI, and the NLMO/NPA bond order, the values are close to 1 and distinctly less than 1.5. Relatively larger contributions of ionic interactions were found in the RGBI as well as NBO bond orders. In comparison with the "C=C" bond in **3a**, the double-bond character is increased in the case of **4a**. This fact is also reflected in ^{13}C NMR data in solution. However, our analyses using the XWR wavefunctions reveal that the C2 atom in **4a** still remains a partial carbanion. Such a borderline case between the π-bonding resonance structure and the charge-separated one makes it difficult to represent the molecular structure of **4a** by a single simple Lewis-structure-based drawing; only a weighted combination of **4a-A** and **4a-B** can correctly describe the bonding situation.

Experimental Section

General synthetic procedure of 3 and 4: Carbodiimide (0.55 mmol) was added to a solution of Tf_2CH_2 **2** (0.50 mmol) in CH_2Cl_2 or $ClCH_2CH_2Cl$ (9.0 mL). After being stirred at room temperature or 50 °C, the reaction mixture was evaporated and purified by column chromatography on neutral silica gel to give the desired push–pull ethylenes.

Acknowledgements

This work was partially supported by a Grant-in-Aid for Scientific Research (C) (No. 17K08224), the Hoansha Foundation, and the German Research Foundation (project GR 4451/1-1). F.K. thanks the German Academic Scholarship Foundation for financial support of his PhD project. We thank Prof. Gerd-Volker Röschenthaler (Jacobs University, Bremen), Dr. Maksym V. Ponomarenko (Jacobs University, Bremen), and Malte Fugel (Universität Bremen) for helpful discussions. We also thank Prof. Manabu Kiguchi and Dr. Satoshi Kaneko (Tokyo Institute of Technology) for help with Raman spectroscopy.

Conflict of interest

The authors declare no conflict of interest.

Keywords: alkenes · complementary bonding analysis · carbanions · quantum crystallography · zwitterions

How to cite: *Angew. Chem. Int. Ed.* **2019**, *58*, 8839–8844
Angew. Chem. **2019**, *131*, 8931–8936

---

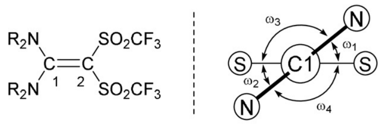
[1] We are aware of the fact that, strictly speaking, we cannot reveal the nature of the C=C bond, but we only use models of electronic structure to describe chemical bonding and therefore approximate the nature of the chemical bond. For more insights into the difference between the description of bonding using models and the causes and origins of chemical bonding, we refer to work by Ruedenberg and Kutzelnigg: a) K. Ruedenberg, *Rev. Mod. Phys.* **1962**, *34*, 326; b) K. Ruedenberg, M. W. Schmidt, *J. Comput. Chem.* **2007**, *28*, 391; c) "The Physical Origin of the Chemical Bond": W. Kutzelnigg in *The Concept of the Chemical Bond* (Ed.: Z. B. Maksic), Springer, Heidelberg, **1990**.

[2] For reviews, see: a) J. Sandström, *Top. Stereochem.* **1983**, *14*, 83; b) E. Kleinpeter, *J. Serb. Chem. Soc.* **2006**, *71*, 1.

[3] a) For reviews, see: P. U. Biedermann, J. J. Stezowski, I. Agranat, *Eur. J. Org. Chem.* **2001**, 15234; b) C. de la Harpe, W. E. van Dorp, *Ber. Dtsch. Chem. Ges.* **1875**, *8*, 1048; c) H. Takezawa, T. Murase, M. Fujita, *J. Am. Chem. Soc.* **2012**, *134*, 17420.

[4] a) R. Destro, U. Cosentino, G. Moro, E. Ortoleva, T. Pilati, *J. Mol. Struct.* **1989**, *212*, 97; b) A. Forni, R. Destro, *Chem. Eur. J.* **2003**, *9*, 5528.

[5] We used structural parameters τ, χ_{C1}, and χ_{C2}, which were originally proposed for twisted amide chemistry: a) F. K. Winkler, J. D. Dunitz, *J. Mol. Biol.* **1971**, *59*, 169; b) S. Yamada, *Angew. Chem. Int. Ed. Engl.* **1993**, *32*, 1083; *Angew. Chem.* **1993**, *105*, 1128.



$$\tau = |1/2\,(\omega_1 + \omega_2)|$$
$$\chi_{C1} = \omega_4 - \omega_1 + \pi = -\omega_3 + \omega_2 - \pi$$
$$\chi_{C2} = \omega_3 - \omega_1 + \pi = -\omega_4 + \omega_2 - \pi$$

[6] H. Yanai, H. Ogura, H. Fukaya, A. Kotani, F. Kusu, T. Taguchi, *Chem. Eur. J.* **2011**, *17*, 11747.

[7] a) R. J. Koshar, R. A. Mitsch, *J. Org. Chem.* **1973**, *38*, 3358; b) A. Hasegawa, T. Ishikawa, K. Ishihara, H. Yamamoto, *Bull. Chem. Soc. Jpn.* **2005**, *78*, 1401; c) H. Yanai, S. Egawa, K. Yamada, J. Ono, M. Aoki, T. Matsumoto, T. Taguchi, *Asian J. Org. Chem.* **2014**, *3*, 556; d) D. Höfler, R. Goddard, J. B. Lingnau, N. Nöthling, B. List, *Angew. Chem. Int. Ed.* **2018**, *57*, 8326; *Angew. Chem.* **2018**, *130*, 8459.

[8] a) H. Yanai, T. Yoshino, M. Fujita, H. Fukaya, A. Kotani, F. Kusu, T. Taguchi, *Angew. Chem. Int. Ed.* **2013**, *52*, 1560; *Angew. Chem.* **2013**, *125*, 1600; b) H. Yanai, Y. Takahashi, H. Fukaya, Y. Dobashi, T. Matsumoto, *Chem. Commun.* **2013**, *49*, 10091; c) H. Yanai, P. Almendros, S. Takahashi, C. Lázaro-Milla, B. Alcaide, T. Matsumoto, *Chem. Asian J.* **2018**, *13*, 1956.

[9] For reviews, see: a) H. Yanai, T. Taguchi, *J. Fluorine Chem.* **2015**, *174*, 108; b) T. Akiyama, K. Mori, *Chem. Rev.* **2015**, *115*, 9277; c) H. Yamamoto, D. Nakashima in *Acid Catalysis in Modern Organic Synthesis*, *Vol. 1* (Eds.: H. Yamamoto, K. Ishihara), Wiley-VCH, Weinheim, **2008**, pp. 35–62.

[10] For selected examples, see: a) K. Ishihara, A. Hasegawa, H. Yamamoto, *Angew. Chem. Int. Ed.* **2001**, *40*, 4077; *Angew. Chem.* **2001**, *113*, 4201; b) A. Hasegawa, Y. Naganawa, M. Fushimi, K. Ishihara, H. Yamamoto, *Org. Lett.* **2006**, *8*, 3175; c) A. Takahashi, H. Yanai, T. Taguchi, *Chem. Commun.* **2008**, 2385; d) H. Yanai, O. Kobayashi, K. Takada, T. Isono, T. Satoh, T. Matsumoto, *Chem. Commun.* **2016**, *52*, 3280; e) D. Höfler, M. van Gemmeren, P. Wedemann, K. Kaupmees, I. Leito, M. Leutzsch, J. Lingnau, B. List, *Angew. Chem. Int. Ed.* **2017**, *56*, 1411; *Angew. Chem.* **2017**, *129*, 1433; f) H. Yanai, R. Takahashi, Y. Takahashi, A. Kotani, H. Hakamata, T. Matsumoto, *Chem. Eur. J.* **2017**, *23*, 8203.

[11] M. Hanack, J. Hackenberg, O. Menke, L. R. Subramanian, R. Schlichenmaier, *Synthesis* **1994**, 249.

[12] H. Akutsu, K. Nakashima, H. Yanai, A. Kotani, S. Hirashima, T. Yamamoto, R. Takahashi, A. Yoshida, Y. Koseki, H. Hakamata, T. Matsumoto, T. Miura, *Synlett* **2017**, *28*, 1363.

[13] A. Z.-Q. Khan, J. Sandstram, *J. Org. Chem.* **2015**, *80*, 90.
[14] J. F. Nicoud, R. J. Twieg in *Nonlinear Optical Properties of Organic Molecules and Crystals* (Eds.: D. S. Chemla, J. Zyss), Academic Press, New York, **1987**, pp. 227–296.
[15] a) D. Jayatilaka, B. Dittrich, *Acta Crystallogr. Sect. A* **2008**, *64*, 383; b) S. C. Capelli, H.-B. Bürgi, B. Dittrich, S. Grabowsky, D. Jayatilaka, *IUCrJ* **2014**, *1*, 361.
[16] M. Woińska, S. Grabowsky, P. M. Dominiak, K. Woźniak, D. Jayatilaka, *Sci. Adv.* **2016**, *2*, e1600192.
[17] F. H. Allen, O. Kennard, D. G. Watson, L. Brammer, A. G. Orpen, R. Taylor, *J. Chem. Soc. Perkin Trans. 2* **1987**, S1.
[18] a) D. Jayatilaka, *Phys. Rev. Lett.* **1998**, *80*, 798; b) D. Jayatilaka, D. J. Grimwood, *Acta Crystallogr. Sect. A* **2001**, *57*, 76.
[19] a) M. Woińska, D. Jayatilaka, B. Dittrich, R. Flaig, P. Luger, K. Woźniak, P. M. Dominiak, S. Grabowsky, *ChemPhysChem* **2017**, *18*, 3334; b) S. Grabowsky, P. Luger, J. Buschmann, T. Schneider, T. Schirmeister, A. N. Sobolev, D. Jayatilaka, *Angew. Chem. Int. Ed.* **2012**, *51*, 6776; *Angew. Chem.* **2012**, *124*, 6880.
[20] S. Grabowsky, A. Genoni, H.-B. Bürgi, *Chem. Sci.* **2017**, *8*, 4159.
[21] M. D. Gould, C. Taylor, S. K. Wolff, G. S. Chandler, D. Jayatilaka, *Theor. Chem. Acc.* **2008**, *119*, 275.
[22] a) *The Quantum Theory of Atoms in Molecules* (Eds.: C. F. Matta, R. J. Boyd), Wiley-VCH, Weinheim, **2007**; b) R. F. W. Bader, *Atoms in Molecules: A Quantum Theory*, Oxford University Press, Oxford, **1990**.
[23] a) M. Kohout, *Int. J. Quantum Chem.* **2004**, *97*, 651; b) M. Kohout, F. R. Wagner, Y. Grin, *Theor. Chem. Acc.* **2008**, *119*, 413.
[24] S. Raub, G. Jansen, *Theor. Chem. Acc.* **2001**, *106*, 223.
[25] F. Weinhold, C. R. Landis, *Valency and Bonding: A Natural Bond Orbital Donor-Acceptor Perspective*, Cambridge University Press, Cambridge, **2004**.
[26] E. D. Glendening, F. Weinhold, *J. Comput. Chem.* **1998**, *19*, 593.
[27] a) R. F. W. Bader, M. E. Stephens, *J. Am. Chem. Soc.* **1975**, *97*, 7391; b) X. Fradera, M. A. Austen, R. F. W. Bader, *J. Phys. Chem. A* **1999**, *103*, 304; c) X. Fradera, J. Poater, S. Simon, M. Duran, M. Solà, *Theor. Chem. Acc.* **2002**, *108*, 214.
[28] The calculation of the DI at the DFT level does not include correlation effects. See: J. Poater, M. Solà, M. Duran, X. Fradera, *Theor. Chem. Acc.* **2002**, *107*, 362.
[29] e.g., S. Wojtulewski, S. J. Grabowski, *J. Mol. Struct. (Theochem)* **2003**, *621*, 285.
[30] T. Ziegler, A. Rauk, *Theor. Chim. Acta* **1977**, *46*, 1.

Manuscript received: April 5, 2019
Accepted manuscript online: April 26, 2019
Version of record online: May 23, 2019

4.3 *Publication:* Proximity Enforced Agostic Interactions Involving Closed-Shell Coinage Metal Ions

The publication called "Proximity Enforced Agostic Interactions Involving Closed-Shell Coinage Metal Ions" has contributions through my program *cuQCrT*. The calculation of very precise grids for the analysis of ELI and $\triangle\rho$ was done using the corresponding implementation in *cuQCrT*. The user interface was documented and made accessible for other users and example files written to be used on this specific case.

Figure 2 in this publication shows how important it is to be able to calculate very accurate volumetric datasets to understand the non-classic interactions inside a molecule or complex and how it can even show interactions of closed-shell metal ions like Cu^+, Zn^+, Ag^+ or Au^+. Also, this example shows how important it is to describe the electron density of these transition metal accurately, which is additional proof that an improvement of the evaluation of densities for a general applicability of HAR was necessary, as performed in *NoSpherA2*.

My contribution to this publication is the TOC Figure, compilation of *cuQCrT* to efficiently calculate the grids necessary, recalculation of wavefunctions without ECP on final geometries to make these detailed pictures and assistance in setting up necessary calculations in *cuQCrT*.

Reprinted with permission from *Inorganic Chemistry* **2019**, *58*, 24, 16372-16378. Copyright © 2019 American Chemical Society.

Inorganic Chemistry
Article
Cite This: Inorg. Chem. 2019, 58, 16372−16378
pubs.acs.org/IC

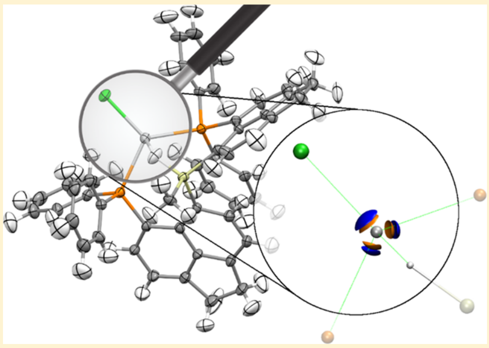
Proximity Enforced Agostic Interactions Involving Closed-Shell Coinage Metal Ions

Emanuel Hupf,[*,†,⊥] Lorraine A. Malaspina,[†] Sebastian Holsten,[†] Florian Kleemiss,[†,∥] Alison J. Edwards,[‡] Jason R. Price,[§] Valeri Kozich,[#] Karsten Heyne,[#] Stefan Mebs,[*,#] Simon Grabowsky,[*,†,∥] and Jens Beckmann[*,†]

[†]Institut für Anorganische Chemie und Kristallographie, Universität Bremen, Leobener Straße 3 & 7, 28359 Bremen, Germany
[⊥]Department of Chemistry, University of Alberta, 11227 Saskatchewan Dr., Edmonton, Alberta T6G 2G2, Canada
[∥]Departement für Chemie und Biochemie, Universität Bern, Freiestrasse 3, 3012 Bern, Switzerland
[‡]Australian Centre for Neutron Scattering, ANSTO, New Illawarra Road, Lucas Heights, NSW 2234, Australia
[§]MX Beamlines, Australian Synchrotron, ANSTO, 800 Blackburn Road, Clayton, Victoria 3168, Australia
[#]Institut für Experimentalphysik, Freie Universität Berlin, Arnimallee 14, 14195 Berlin, Germany

Supporting Information

ABSTRACT: A proximity enforcing diarylsilane ligand is reported, which gives rise to unusual Si−H···M interactions with the d^{10} metal ions Cu^+ and Ag^+ upon complexation. These interactions are studied in detail both experimentally and computationally and can be classified to be weakly agostic in nature for the Si−H···Cu interaction. The Si−H···Ag interaction has more signatures of an electrostatic contact.



INTRODUCTION

Agostic interactions are originally defined as interactions between a C−H bond and a metal atom. They are partially covalent 2e−3c C−H···M interactions, in contrast to predominantly electrostatic hydrogen bonds.[1] The H···M interaction may activate the C−H bond and, hence, plays an important role in fundamental organometallic transformations.[2] The C−H activation process is governed by the interaction of a formally hydridic hydrogen atom with a local Lewis acidic site in the structured outer core of the metal atom, identified in terms of a charge depletion.[3] This charge depletion is analogous to a vacant d-orbital in the molecular orbital picture.[4] These charge depletion zones can most easily be formed in transition metals with partially unfilled d-shells[5] but have also been observed in d^0 metals.[6] Formation of a charge depletion in putatively spherical saturated d^{10} metals lacking vacant d-orbitals remains a challenge, which makes agostic interactions involving d^{10} coinage metal ions rare and weak.[7] Stronger agostic interactions with coinage metals can only be formed in higher oxidation states,[8] or when the donor group is a Si−H instead of a C−H bond, for which there is, to the best of our knowledge, only a single claim of a complex comprising a Si−H···Cu (d^{10}) interaction.[9]

Nonclassical Si−H···M interactions include σ-complexes, asymmetric and symmetric oxidative addition products as well as silyl hydrides.[10] In addition, Si−H···M interactions can also show the general signatures of agostic interactions (although the original definition referred to C−H···M), because in the Si−H bond, the charge separation and hence the hydridic character is more pronounced than in C−H bonds, so that Si−H bond activation with nonsaturated transition metals for oxidative addition reactions such as hydrosilylations is common.[11] A Si−H···M agostic interaction can be understood as the intramolecular analogue of a σ-complex describable by the Dewar−Chatt−Duncanson model[12] involving metal−ligand π back-donation.[13,14] The degree of Si−H activation correlates with the degree of π back-bonding, so that complexes of varying strength can be interpreted as snapshots along an asymmetric oxidative addition reaction coordinate.[15]

Criteria for the identification and characterization of Si−H···M agostic interactions involve: a significant Si−H elongation (Si−H > 1.6 Å)[16] in conjunction with an IR red shift[13] and a decreased absolute value of the (negative) NMR coupling constant ^1J(Si−H)[10,17] compared to an Si−H bond not

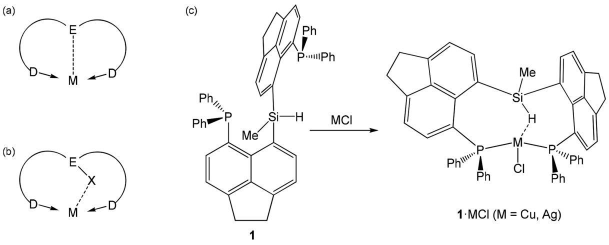
Received: June 21, 2019
Published: November 20, 2019

DOI: 10.1021/acs.inorgchem.9b01845
Inorg. Chem. 2019, 58, 16372−16378

involved in an interaction, decreasing Si−H···M bond angle and increasing Si···M···H angle with increasing strength of the Si−H···M agostic interaction,[10] a topological electron-density pattern indicating the presence of a M-H bond,[18] and the identification of a charge depletion (Lewis acidic site) at the metal ion core.[15,10]

In this study, we investigate Si−H···M (M = d^{10} coinage metal ions Cu^+ and Ag^+) interactions. The Si−H and M groups are brought into close proximity by ligand **1** to ensure a close contact of the Si−H bond to the Cu^+ and Ag^+ ions due to geometric constraints (Scheme 1). We therefore refer to **1** as a *proximity enforcing ligand* (PEL). The Si−H···M interaction in **1**·MCl is, hence, an *enforced proximity interaction* (EPI),[19] whereby the nature of the interaction is unknown. It could be repulsive or attractive, most likely weak, which is to be clarified in this study. It should be noted that the EPIs present in **1**·MCl are caused by the unique nature of the PEL **1**, which differentiates it from solely chelating ligands.

Scheme 1. Principle of Enforced Proximity Interactions (EPIs) Utilizing Proximity Enforcing Ligands (PELs) via (a) Direct Interaction between E and M or (b) Interaction between X and M[a]



[a]In this study, the reaction of PEL **1** with MCl (c) gives rise to the formation of complexes **1**·MCl possessing M···H interactions of type (b).

RESULTS AND DISCUSSION

PEL **1** was prepared in a single step from 5-diphenylphosphinoacenaphth-6-yl lithium with methyldichlorosilane. Addition of CuCl or AgCl provided the complexes **1**·CuCl and **1**·AgCl in yields of 66% and 90%, respectively.

NMR spectroscopy facilitates quantification of the effects caused by the intramolecular Si−H···M interactions. The ^1H NMR chemical shift of the Si−H bond marginally changes in the case of **1**·CuCl (δ = 6.95 ppm) but is significantly downfield shifted in **1**·AgCl (δ = 7.85 ppm) compared to the free ligand **1** (δ = 7.13 ppm).[20] The ^{29}Si NMR chemical shifts of **1**, **1**·CuCl, and **1**·AgCl are very similar, with the chemical shift of **1**·CuCl (δ = −29.0 ppm) being slightly upfield shifted compared to **1** (δ = −23.3 ppm) and **1**·AgCl (−20.6 ppm).

Upon complexation with CuCl, the $^1J(^1H-^{29}Si)$ coupling constant decreases significantly from an absolute value of 228.6 Hz in **1** (theoretical value: −218.1 Hz at B3PW91/6-311+G(2df,p); −216.6 Hz at B3PW91/DGDZVP [all-electron calculations based on the optimized geometries *vide infra*]) to an absolute value of 161.6 Hz in **1**·CuCl (theoretical value: −166.1 Hz at B3PW91/DGDZVP and −159.0 Hz with the inclusion of empirical dispersion), being in the same range as the related $[Ph_2P(o\text{-}C_6H_4)]_2SiMeH\cdot CuCl$,[9] indicating a significant weakening of the Si−H bond. The same trend, although smaller, is observed for **1**·AgCl, showing an absolute value of the $^1J(^1H-^{29}Si)$ coupling constant of 185.8 Hz (theoretical value: −184.8 at B3PW91/DGDZVP and −188.1 Hz with dispersion).[21]

IR spectroscopy further confirms this trend as a decrease of the wavenumber of the Si−H stretching vibration from $\tilde{\nu}_{SiH}$ = 2173 cm^{-1} in **1** to 2118 cm^{-1} (**1**·AgCl) and 1932 cm^{-1} (**1**·CuCl), depicted in Figure S22, left. Theoretically calculated values at B3PW91/6-311+G(2df,p) level of theory (see geometry optimization below) are $\tilde{\nu}_{SiH}$ = 2169 cm^{-1} (**1**), 2089 cm^{-1} (**1**·AgCl), and 1984 cm^{-1} (**1**·CuCl). In a local mode picture, this frequency shift demonstrates the Si−H bond in the metal complexes to be stretched and weakened because of the Si−H···M interaction, which is more pronounced in **1**·CuCl than in **1**·AgCl.

This is corroborated by the measured excited state decay and ground state bleaching recovery times of the Si−H stretching vibrations. Upon excitation of the Si−H stretching vibration in **1** energy relaxes with 7(1) ps (see Figure S22) to lower frequency modes inconsiderably coupled to the Si−H stretching vibration, as visible from the vanishing signal. Excitation of lower frequency modes with considerable coupling to the Si−H stretching mode would result in spectral shifting of the Si−H absorption band after excited state decay. Si−H stretch excitation in **1**·AgCl displays a very fast decay of the vibrational energy to lower-frequency modes of about 0.5 to 0.8 ps (see Figure S22), explained by either a stronger coupling to the same internal coordinates or by coupling to new coordinates established by the Si−H···Ag interaction. This effect is similar but stronger in **1**·CuCl. Si−H stretch excitation results in an excited state decay faster than the system response of 0.5 ps. Thus, the excitation energy is efficiently redistributed to lower-frequency vibrations, coupled to the Si−H stretching vibration. At least some of these lower-frequency vibrations do couple considerably to the Si−H stretching vibration upon excitation, since the bleaching signal recovers on a longer time scale of 1.6(5) ps. The frequency position of the Si−H stretching vibration in **1**·AgCl and **1**·CuCl is reduced by 2.6% and 12.5%, respectively (theoretical values 7.4% and 16.4%).

Comparison of the absorption frequencies with bond lengths discussed in the next section shows a linear decrease of the frequencies with increasing Si−H bond lengths (see Figure S24), similar to the behavior of the hydrogen-bonded O−H stretching vibration.[22] Time-resolved femtosecond infrared pump−probe spectroscopy was performed with home-built IR optical parametric amplifiers for pump and probe pulses (see the Supporting Information for more details).[23]

These vibrational dynamics experiments indicate only a weak anharmonic coupling of the Si−H stretching vibration to other vibrational modes in **1**, which means that the isolated harmonic stretching vibration can indeed be used as a measure of the agostic interaction. In **1**·AgCl, this anharmonic coupling is slightly stronger, but it is significantly stronger in **1**·CuCl, which is an indication of a more pronounced interaction of the Si−H group with the metal atom.

For the accurate and precise localization of the hydrogen atoms, routine X-ray diffraction experiments were supported by neutron-diffraction studies of **1**, **1**·MCl (M = Cu, Ag), and triphenylsilane, Ph_3SiH. The final geometries of **1** and **1**·MCl derived from the single-crystal neutron diffraction experiments are shown in Figure 1. In Ph_3SiH, the Si−H bond is 1.494(6) Å and involved in a short (2.290(9) Å) $H(\delta^+)\cdots H(\delta^-)$ intermolecular interaction. In PEL **1**, where no intermolecular interactions involving the Si−H bond are

DOI: 10.1021/acs.inorgchem.9b01845
Inorg. Chem. 2019, 58, 16372−16378

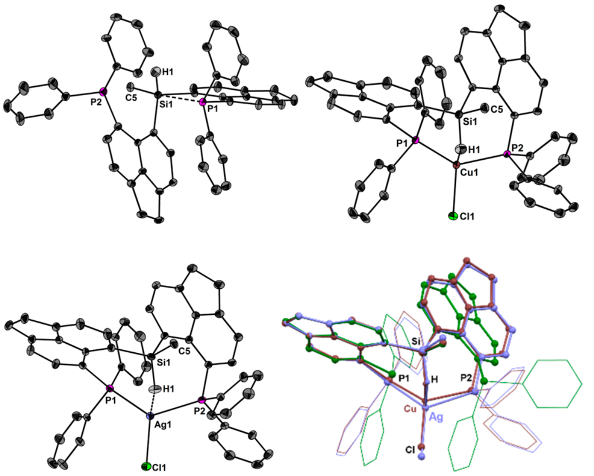


Figure 1. Molecular structures of **1** (top left), **1**·CuCl (top right), and **1**·AgCl (bottom left) established by neutron diffraction showing 30% probability ellipsoids and the crystallographic numbering scheme. A superimposed view of **1** (green), **1**·CuCl (red), and **1**·AgCl (blue) are shown on the bottom right.

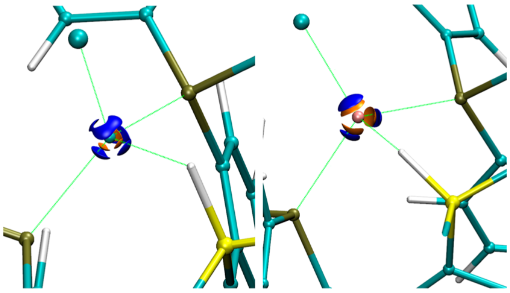


Figure 2. Isosurface representations of ELI-D (blue) and Laplacian of ED (orange) of **1**·CuCl (left) and **1**·AgCl (right) around the metal atoms. Isovalues for Cu: −185.0 e Å^{-5} (Laplacian) and 1.677 (ELI-D). Isovalues for Ag: −43.8 e Å^{-5} (Laplacian) and 1.720 (ELI-D).

present, the bond is shortened to 1.484(7) Å. Relative to the situation in **1**, the Si−H bond is elongated significantly to 1.509(7) Å in **1**·CuCl and is within a single standard deviation in **1**·AgCl (1.492(10) Å), consistent with results from IR and NMR spectroscopy.

The Si···Cu and Cu···H distances in **1**·CuCl are 2.812(5) and 1.775(7) Å. Both values are significantly below the sum of the van-der-Waals radii of 4.10 Å for Si···Cu and 3.10 Å for Cu···H.[24] The same situation, although less pronounced, is observed for **1**·AgCl. The Si···Ag distance is 3.157(9) Å, and the Ag···H distance is 2.090(11) Å, which are also below the van-der-Waals radii of 4.20 Å for Si···Ag and 3.20 Å for Ag···H.[24] The Si−H···M angles are significantly bent with values of 117.6(4)° (**1**·CuCl) and 122.7(6)° (**1**·AgCl), respectively, indicative of weak agostic interactions.

Geometries were fully optimized at the B3PW91/6-311+G-(2df,p) level of theory with an effective core potential (ECP10/28MDF) and a corresponding cc-pVTZ basis set for both Cu and Ag atoms. The resulting optimized geometries are very similar to the experimental neutron-diffraction models. Based on the optimized geometries, real-space bonding indicators (RSBIs) were analyzed. These comprise topological dissection of the calculated electron density (ED) according to the Atoms-In-Molecules (AIM)[25] approach and of the pair density according to the electron localizability indicator (ELI-D)[26] approach, complemented by the noncovalent interaction (NCI) index[27] surface representations (Figures 2−4). Furthermore, compliance constants (the inverse of relaxed force constants) and coupling compliance constants have been determined applying the generalized compliance constants method.[28] This combination of different theoretical approaches is used to accurately characterize the bonding situation, which is of importance especially for weak interactions.[29]

One of the most important criteria for the occurrence of an agostic interaction is the existence of local charge depletion zones at the metal core that normally indicate vacant d-orbitals.[15,10] These charge depletions at the metal ions are clearly present in both compounds induced by the ligand field in the PEL, although the Cu$^+$ and Ag$^+$ ions are formally of closed-shell d^{10} configuration (Figure 2). Natural population analysis[30] reveals a small deviation from the ideal d^{10} population in **1**·CuCl (d$^{9.84}$) and **1**·AgCl (d$^{9.89}$) (B3PW91/DGDZVP level of theory, all-electron calculations based on the optimized geometries, more details in Table S6). Hence, charge reorganization within the d-shell accommodates agostic interactions in both compounds, with the respective Si−H group pointing toward a depletion zone. This situation is more pronounced in the Ag atom with its larger, more polarizable shell. Clearly structured Ag−P and Ag−Cl valence shell charge concentrations indicate covalent contributions to the bonding. The smaller and hence less polarizable Cu ion shows the same structure for the Cu−P bonds, but a more diffuse charge distribution otherwise.

AIM atomic charges of about −0.7e for the hydrogen atoms in both **1**·CuCl and **1**·AgCl show the significant hydridic character of these hydrogen atoms which facilitates agostic bonding. The spatial arrangement of the metal atoms in **1**·CuCl and **1**·AgCl is best described as distorted tetrahedral taking into account the stable bond path to the hydrogen atom (Figures 3a and 4a). The AIM bonding properties (Table S4) show that the H···Cu interaction has about twice as much electron density accumulated in its bond critical point (bcp) in comparison to the H···Ag interaction (0.43 vs 0.26 e Å^{-3}). It also has a significantly higher negative total energy density at the bcp (−0.26 vs −0.09 au). This means that the H···Cu interaction is stronger and more covalent, whereas the H···Ag interaction is weaker and electrostatically dominated. This interpretation is supported by the ELI-D analysis (Figures 3c,d and 4c,d). Both Si−H basins in **1**·CuCl and **1**·AgCl share a boundary with the metal atoms, but the basin is flat along the H···Ag axis and curved along the H···Cu axis.

In the NCI of **1**·CuCl, a ring-shaped and mixed-colored NCI surface (Figures 3b and 4b) is formed along the H−Cu axis which is typical for a polar covalent interaction similar to the Cu/Ag−P/Cl interactions and in line with the interpretation of the Cu···H interaction to be agostic. In **1**·AgCl, a disc-shaped and blue-colored basin (fused with a flat and green-colored interligand region) is formed along the H−Ag axis pointing toward a predominantly electrostatic H···Ag contact. More details of the RSBI analysis are given in the Supporting Information.

Determination of the compliance matrix is a powerful tool in unraveling weak agostic interactions.[28b] Parts of the compliance matrix are given in Table S7. The compliance constant C_{ii} for the Si−H bond increases in the order C_{SiH} = 0.351 Å

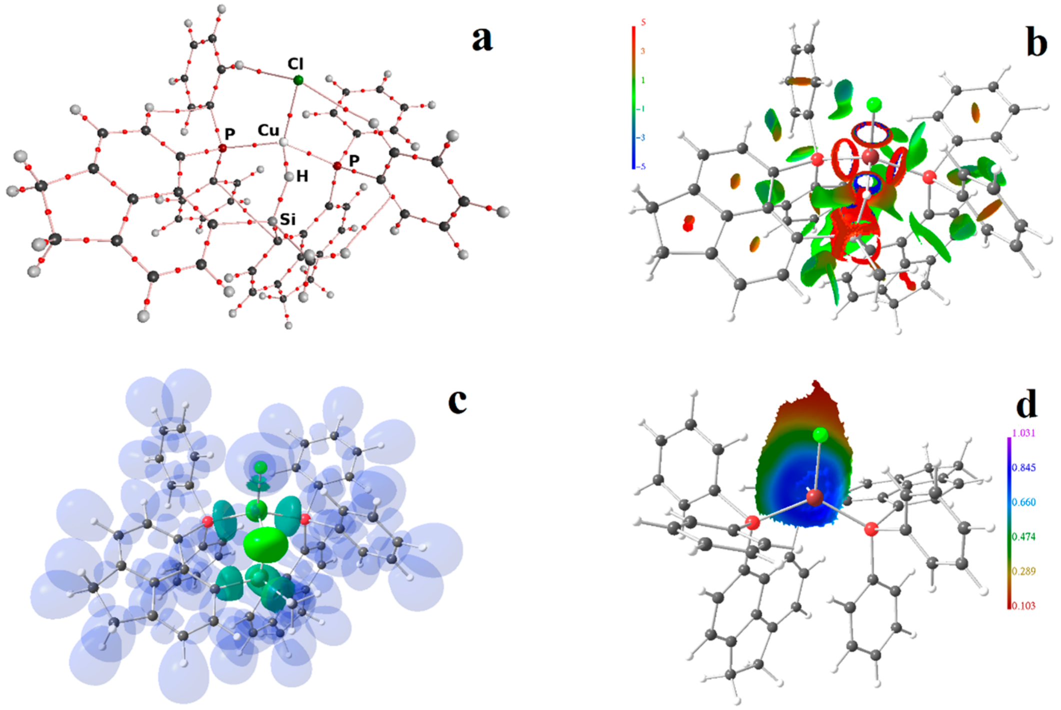


Figure 3. RSBI analysis of **1**·CuCl. (a) AIM bond paths motif with bond critical points (bcps, red spheres). (b) NCI *iso*-surface at $s(r) = 0.5$ color coded with $\text{sign}(\lambda_2)\rho$ in a.u. Blue surfaces refer to attractive forces, and red to repulsive forces. Green indicates weak interactions. (c) ELI-D localization domain representation at an *iso*-value of 1.3. (d) ELI-D distribution (unitless) color-coded onto the (Si−)H ELI-D basin.

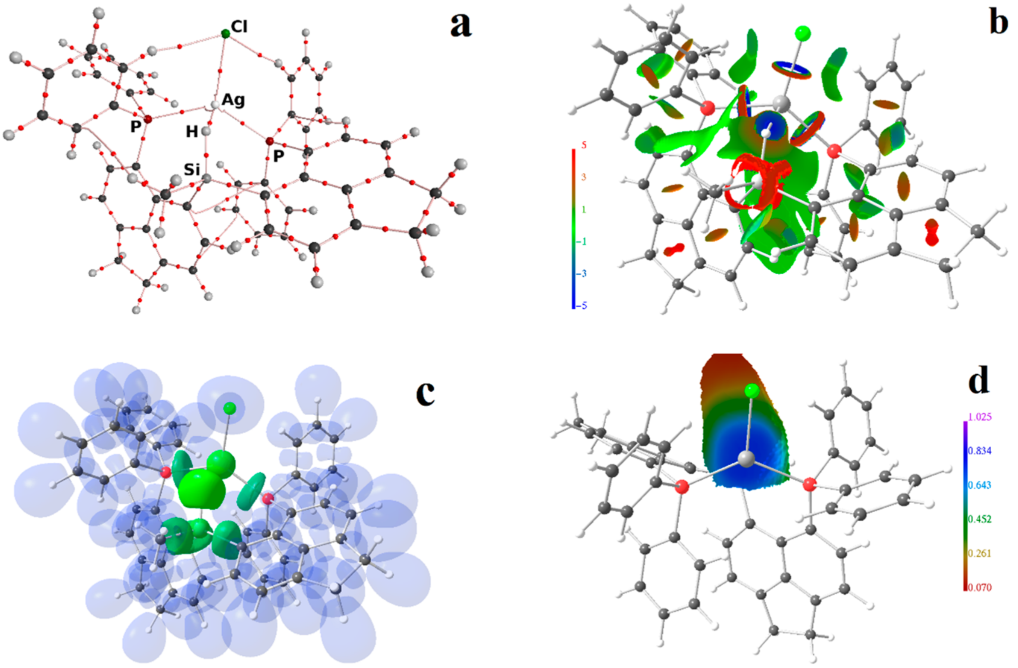


Figure 4. RSBI analysis of **1**·AgCl. (a) AIM bond paths motif with bond critical points (bcps, red spheres). (b) NCI *iso*-surface at $s(r) = 0.5$ color coded with $\text{sign}(\lambda_2)\rho$ in a.u. Blue surfaces refer to attractive forces and red to repulsive forces. Green indicates weak interactions. (c) ELI-D localization domain representation at an *iso*-value of 1.3. (d) ELI-D distribution (unitless) color-coded onto the (Si-)H ELI-D basin.

mdyn^{-1} (**1**) < 0.389 Å mdyn^{-1} (**1**·AgCl) < 0.475 Å mdyn^{-1} (**1**·CuCl). As larger numerical values are indicative of weaker bonds, the Si−H bond is considerable weakened in the Cu-complex. In line, the Cu−H interaction shows a significantly smaller compliance constant (C_{CuH} = 2.828 Å mdyn^{-1}) than the respective Ag−H interaction (C_{AgH} = 6.239 Å mdyn^{-1}) pointing to a stronger M···H interaction in **1**·CuCl. Furthermore, the value of C_{SiCu} is significantly larger (6.102

DOI: 10.1021/acs.inorgchem.9b01845
Inorg. Chem. 2019, 58, 16372−16378

Å mdyn^{-1}) than the C_{SiH} or C_{CuH} values. In contrast, for **1**·AgCl, the C_{SiAg} value of 6.142 Å mdyn^{-1} is comparable to the C_{AgH} value, which further shows the stronger agostic interaction in **1**·CuCl.

A study of the compliance coupling constants (see Table S7) reveals substantial coupling between all three fragments of interest, namely Si−H, M−H, and Si−M (M = Cu, Ag). As the compliance coupling constants are directly proportional to the interaction of the coupling coordinates,[28b] all three fragments contribute substantially to the overall lowering of the total energy.

In order to study the effect of London dispersion on the Si−H···M interactions, empirical dispersion effects were included using Grimme's GD3BJ parameters.[31] The reoptimized geometries give rise to longer Si−H and substantially shorter H···M bond distances compared to the geometries without inclusion of dispersion effects (see the Supporting Information for details). However, RSBI analysis reveals very similar properties in case of **1**·AgCl. Interestingly, for **1**·CuCl, a small increase in pronounced Si−H···Cu interactions is detected, e.g. the electron density and Laplacian at the bcp for the Si−H bond is decreasing along with a simultaneous increase in electron density and Laplacian values at the bcp for the H···Cu interaction (Table S4 and S5, Figure S26 and S27). In contrast, the effect of dispersion on the compliance matrix is quite substantial, especially on the M−H interaction. The value of C_{SiH} in **1**·CuCl increases to 0.670 Å mdyn^{-1}, clearly demonstrating the weakening of the Si−H bond, especially when compared to the C_{SiH} value of **1**·AgCl, which remains almost unchanged upon inclusion of empirical dispersion corrections (0.395 Å mdyn^{-1}). Simultaneously, the C_{MH} and C_{SiM} values decrease by 31−33% and 13−20%, respectively (Table S7). Although the overall picture of the Si−H···M interactions can be understood without the inclusion of London dispersion effects, the explicit treatment of London dispersion reveals a more detailed and pronounced description of the interaction and clearly shows the stronger H···Cu interaction in **1**·CuCl compared to the **1**·AgCl congener. Hence, stabilization of interactions by London dispersion is part of the working principle of the PELs.

■ CONCLUSIONS

The Si−H···Cu/Ag interactions present in **1**·MCl do in fact describe attractive forces mediated by the proximity enforcing ligands. The stretching frequencies, the vibrational dynamics, and the geometric, NMR, and electronic parameters discussed here are signs of agostic interactions between the Si−H and metal groups, which is more strongly pronounced in the Si−H···Cu (predominantly covalent contributions) than in the Si−H···Ag complex (predominantly electrostatic contributions). Relative to the criteria listed in the introduction, the Si−H···Cu agostic interaction is weak; sometimes above, sometimes below the borderline for such interactions, so that the term pregostic interaction might be suitable,[32] whereas the Si−H···Ag interaction is better described as an electrostatic contact with some signatures of an agostic interaction.

We have shown here that in a suitable predominantly covalent ligand environment, closed-shell metal ion cores can be polarized breaking the spherical symmetry of the d^{10} shell, so that agostic interactions can be formed. To establish this suitable environment, we synthesized and used a proximity enforcing ligand (PEL). This PEL stabilizes interactions that would not form under normal circumstances. It has been reported that pincer ligands fail to fulfill the same task for coinage metal agostic interactions.[33] Therefore, we propose that the principle of proximity enforcing ligands can be more broadly applied to stabilize rare and unusual bonding situations.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.inorgchem.9b01845.

Experimental details: Synthesis, spectroscopy including NMR spectra, vibrational dynamics, and X-ray and neutron diffraction. **Computational details**: Methodology and software, detailed RSBI and NBO analyses, optimized coordinates (PDF)

3D rotatable models (XYZ)

Accession Codes

CCDC 1875890−1875897 contain the supplementary crystallographic data for this paper. These data can be obtained free of charge via www.ccdc.cam.ac.uk/data_request/cif, or by emailing data_request@ccdc.cam.ac.uk, or by contacting The Cambridge Crystallographic Data Centre, 12 Union Road, Cambridge CB2 1EZ, UK; fax: +44 1223 336033.

■ AUTHOR INFORMATION

Corresponding Authors

*E-mail: hupf@uni-bremen.de (E.H.).
*E-mail: stebs@chemie.fu-berlin.de (S.M.).
*E-mail: simon.grabowsky@dcb.unibe.ch (S.G.).
*E-mail: j.beckmann@uni-bremen.de (J.B.).

ORCID

Emanuel Hupf: 0000-0003-0121-7554
Stefan Mebs: 0000-0003-2877-3577
Simon Grabowsky: 0000-0002-3377-9474
Jens Beckmann: 0000-0002-8548-1821

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

S.G. and E.H. thank the German Research Foundation (Deutsche Forschungsgemeinschaft DFG) for fellowships within the schemes GR 4451/1-1 and HU 2512/1-1. S.G. and J.B. thank the DFG furthermore for funding of this research within the projects GR 4451/2-1 and BE 3716/7-1. The neutron diffraction experiments were supported by the Australian Nuclear Science and Technology Organisation (ANSTO) under the Bragg Institute proposal nos. 4766 and 5279. We acknowledge an anonymous referee for valuable discussions of the generalized compliance constants methodology during the reviewing process.

■ REFERENCES

(1) Brookhart, M.; Green, M. L. H.; Parkin, G. Agostic interactions in transition metal compounds. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104*, 6908−6914.

(2) (a) Crabtree, R. H. Transition Metal Complexation of σ Bonds. *Angew. Chem., Int. Ed. Engl.* **1993**, *32*, 789−805. (b) Kubas, G. J. *Metal Dihydrogen and σ-Bond Complexes (Modern Inorganic Chemistry)*; Kluwer Academic/Plenum Publishers: New York, 2001.

(3) (a) Scherer, W.; Herz, V.; Hauf, C. On the Nature of β-Agostic Interactions: A Comparison Between the Molecular Orbital and Charge Density Picture. In *Electron Density and Chemical Bonding I.*

Structure and Bonding; Stalke, D., Ed.; Springer: Berlin, Heidelberg, 2012, *146*, 159−207. (b) Scherer, W.; Wolstenholme, D. J.; Herz, V.; Eickerling, G.; Bruck, A.; Benndorf, P.; Roesky, P. W. On the Nature of Agostic Interactions in Transition-Metal Amido Complexes. *Angew. Chem., Int. Ed.* **2010**, *49*, 2242−2246.

(4) Bader, R. F. W.; Matta, C. F.; Cortes-Guzman, F. Where To Draw the Line in Defining a Molecular Structure. *Organometallics* **2004**, *23*, 6253−6263.

(5) (a) Scherer, W.; Herz, V.; Brück, A.; Hauf, C.; Reiner, F.; Altmannshofer, S.; Leusser, D.; Stalke, D. The Nature of *β*-Agostic Bonding in Late-Transition-Metal Alkyl Complexes. *Angew. Chem., Int. Ed.* **2011**, *50*, 2845−2849. (b) Scherer, W.; Dunbar, A. C.; Barquera-Lozada, J. E.; Schmitz, D.; Eickerling, G.; Kratzert, D.; Stalke, D.; Lanza, A.; Macchi, P.; Casati, N. P. M.; Ebad-Allah, J.; Kuntscher, C. Anagostic Interactions under Pressure: Attractive or Repulsive. *Angew. Chem., Int. Ed.* **2015**, *54*, 2505−2509.

(6) Scherer, W.; McGrady, G. S. Agostic interactions in d^0 metal alkyl complexes. *Angew. Chem., Int. Ed.* **2004**, *43*, 1782−1806.

(7) (a) Ilie, A.; Raţ, C. I.; Scheutzow, S.; Kiske, C.; Lux, K.; Klapötke, T. M.; Silvestru, C.; Karaghiosoff, K. Metallophilic Bonding and Agostic Interactions in Gold(I) and Silver(I) Complexes Bearing a Thiotetrazole Unit. *Inorg. Chem.* **2011**, *50*, 2675−2684 and refs 20 and 21 therein. (b) Corral, I.; Mó, O.; Yáñez, M. Binding energies of Cu$^+$ to saturated and *α*,*β*-unsaturated alkanes, silanes and germanes: The role of agostic interactions. *Int. J. Mass Spectrom.* **2003**, *227*, 401−412. (c) Butler, M. J.; Crimmin, M. R. Magnesium, zinc, aluminium and gallium hydride complex of the transition metals. *Chem. Commun.* **2017**, *53*, 1348−1365.

(8) Rekhroukh, F.; Estveéz, L.; Bijani, C.; Miqueu, K.; Amgoune, A.; Bourissou, D. Experimental and Theoretical Evidence for an Agostic Interaction in a Gold(III) Complex. *Angew. Chem., Int. Ed.* **2016**, *55*, 3414−3418.

(9) Joost, M.; Mallet-Ladeira, S.; Miqueu, K.; Amgoune, A.; Bourissou, D. *σ*-SiH Complexes of Copper: Experimental Evidence and Computational Analysis. *Organometallics* **2013**, *32*, 898−902.

(10) Scherer, W.; Meixner, P.; Batke, K.; Barquera-Lozada, J. E.; Ruhland, K.; Fischer, A.; Eickerling, G.; Eichele, K. *J*(Si, H) Coupling Constants in Nonclassical Transition-Metal Silane Complexes. *Angew. Chem., Int. Ed.* **2016**, *55*, 11673−11677.

(11) Díez-González, S.; Nolan, S. P. Copper, Silver, and Gold Complexes in Hydrosilylation Reactions. *Acc. Chem. Res.* **2008**, *41*, 349−358.

(12) (a) Dewar, M. J. S. A review of the pi-complex theory. *Bull. Chim. Soc. Fr.* **1951**, *18*, C71−C79. (b) Chatt, J.; Duncanson, L. A. Olefin co-ordination compounds. Part III. Infra-red spectra and structure: attempted preparation of acetylene complexes. *J. Chem. Soc.* **1953**, 2939−2947. (c) Kubas, G. J. Metal-Dihydrogen and sigma-bond coordination: The consummate extension of the Dewar-Chatt-Duncanson model for metal-olefin pi-boinding. *J. Organomet. Chem.* **2001**, *635*, 37−68.

(13) Nikonov, G. I. Recent Advances in Nonclassical Interligand Si···H Interactions. *Adv. Organomet. Chem.* **2005**, *53*, 217−309.

(14) (a) Corey, J. Y.; Braddock-Wilking, J. Reactions of Hydrosilanes with Transition-Metal Complexes: Formation of Stable Transition-Metal Silyl Compounds. *Chem. Rev.* **1999**, *99*, 175−292. (b) Kubas, G. J. *Metal Dihydrogen and σ-Bond Complexes: Structure, Theory, and Reactivity*; Springer: Berlin, 2001. (c) Lin, Z. Structural and bonding characteristics in transition metal-silane complexes. *Chem. Soc. Rev.* **2002**, *31*, 239−245. (d) Shimada, S.; Tanaka, M. Group 10 transition-metal complexes with metal-silicon bonds derived from 1,2-disilylbenzenes and bis(2-silylphenyl)silane. *Coord. Chem. Rev.* **2006**, *250*, 991−1011. (e) Corey, J. Y. Reactions of Hydrosilanes with Transition Metal Complexes and Characterization of the Products. *Chem. Rev.* **2011**, *111*, 863−1071. (f) Corey, J. Y. Reactions of Hydrosilanes with Transition Metal Complexes. *Chem. Rev.* **2016**, *116*, 11291−11435. (g) Ramaraj, A.; Reddy, K. H. K.; Keil, H.; Herbst-Irmer, R.; Stalke, D.; Jemmis, E. D.; Jagirdar, B. R. Approach to Sigma Complexes via Displacement of Agostic Interactions: An Experimental and Theoretical Investigation. *Organometallics* **2017**, *36*, 2736−2745.

(15) (a) Scherer, W.; Eickerling, G.; Tafipolsky, M.; McGrady, G. S.; Sirsch, P.; Chatterton, N. P. Elucidation of the bonding in Mn(η^2-SiH) complexes by charge density analysis and T_1 NMR measurements: asymmetric oxidative addition and anomeric effects at silicon. *Chem. Commun.* **2006**, 2986−2988. (b) Scherer, W.; Meixner, P.; Barquera-Lozada, J. E.; Hauf, C.; Obenhuber, A.; Brück, A.; Wolstenholme, D. J.; Ruhland, K.; Leusser, D.; Stalke, D. A Unifying Bonding Concept for Metal Hydrosilane Complexes. *Angew. Chem., Int. Ed.* **2013**, *52*, 6092−6096.

(16) Hauf, C.; Barquera-Lozada, J. E.; Meixner, P.; Eickerling, G.; Altmannshofer, S.; Stalke, D.; Zell, T.; Schmidt, D.; Radius, U.; Scherer, W. Remanent Si-H Interactions in Late Transition Metal Silane Complexes. *Z. Anorg. Allg. Chem.* **2013**, *639*, 1996−2004.

(17) Meixner, P.; Batke, K.; Fischer, A.; Schmitz, D.; Eickerling, G.; Kalter, M.; Ruhland, K.; Eichele, K.; Barquela-Lozada, J. E.; Casati, N. P. M.; Montisci, F.; Macchi, P.; Scherer, W. *J*(Si, H) Coupling Constants of Activated Si-H Bonds. *J. Phys. Chem. A* **2017**, *121*, 7219−7235.

(18) (a) Jabłoński, M. Geometry- and QTAIM-Based Comparison of Intramolecular Charge-Inverted Hydrogen Bonds, M···(H-Si) "Agostic Bond", and M···(*η*2-SiH)*σ* Interactions. *J. Phys. Chem. A* **2015**, *119*, 11384−11396. (b) Jabłoński, M. Comparative study of geometric and QTAIM-based differences between X-H···Y intramolecular charge-inverted hydrogen bonds, M1···(H-X) agostic bonds and M2···(*η*2-XH) *σ* interactions (X = Si, Ge; Y = Al, Ga; M1 = Ti, Co; M2 = Mn, Fe, Cr). *Comput. Theor. Chem.* **2016**, *1096*, 54−65.

(19) The presented systems (**1**·MCl, Scheme 1) are not to be misconstrued as pincer systems (Morales-Morales, D.; Jensen, C. *The Chemistry of Pincer Compounds*, 1st ed.; Elsevier Science, 2007), and PEL **1** is not a pincer ligand, because the metal ion is only strongly coordinated by two sites (here chelating P atoms), enforcing the contact with a third, here the Si−H bond. It is neither to be confused with a *peri*-interaction between the 1,8-positions in a (ace)naphthyl group, for which a similar term (enforced proximity donor) has been used in the literature. Surgenor, B. A.; Bühl, M.; Slawin, A. M. Z.; Woollins, J. D.; Kilian, P. *Angew. Chem., Int. Ed.* **2012**, *51*, 10150−10153.

(20) The chemical shifts of **1**, **1**·CuCl, and **1**·AgCl are obtained from ^{1}H−^{29}Si HSQC NMR measurements with $CDCl_3$ as solvent, due to partial overlap with the aryl region in the 1D ^{1}H NMR; see the Supporting Information.

(21) The Si−*H* chemical shift in the ^{1}H NMR shows a $J(^{107/109}\text{Ag}-^1\text{H})$ coupling constant of 6.8 Hz indicating the presence of an Si−H···Ag interaction in **1**·AgCl. The ^{1}H coupled ^{29}Si NMR spectrum indicates the presence of a $J(^{107/109}\text{Ag}-^{29}\text{Si})$ coupling, which could not unambiguously be distinguished from the $J(^{31}\text{P}-^{29}\text{Si})$ coupling (more details are given in the Supporting Information).

(22) Bratos, S.; Ratajczak, H. Profiles of hydrogen stretching IR bands of molecules with hydrogen bonds: A stochastic theory. II. Strong hydrogen bonds. *J. Chem. Phys.* **1982**, *76*, 77−85.

(23) Stensitzki, T.; Yang, Y.; Kozich, V.; Ahmed, A. A.; Kössl, F.; Kühn, O.; Heyne, K. Acceleration of a ground-state reaction by selective femtosecond-infrared-laser-pulse excitation. *Nat. Chem.* **2018**, *10*, 126−131.

(24) (a) Batsanov, S. Van der Waals Radii of Elements. *Inorg. Mater.* **2001**, *37*, 871−885. (b) Rowland, R. S.; Taylor, R. Intermolecular Nonbonded Contact Distances in Organic Crystal Structures: Comparison with Distances Expected from van der Waals Radii. *J. Phys. Chem.* **1996**, *100*, 7384−7391.

(25) Bader, R. W. F. Atoms in Molecules. *A Quantum Theory*; Cambridge University Press: Oxford U.K., 1991.

(26) Kohout, M. A Measure of Electron Localizability. *Int. J. Quantum Chem.* **2004**, *97*, 651−658.

(27) Johnson, E. R.; Keinan, S.; Mori-Sanchez, P.; Contreras-García, J.; Cohen, A. J.; Yang, W. Revealing Noncovalent Interactions. *J. Am. Chem. Soc.* **2010**, *132*, 6498−6506.

(28) (a) Grunenberg, J.; Goldberg, N. How Strong Is the Ga≡Ga Triple Bond? Theoretical Compliance Matrices as a Probe for Intrinsic Bond Strengths. *J. Am. Chem. Soc.* **2000**, *122*, 6045−6047. (b) von Frantzius, G.; Streubel, R.; Brandhorst, K.; Grunenberg, J. How Strong Is an Agostic Bond? Direct Assessment of Agostic Interactions Using the Generalized Compliance Matrix. *Organometallics* **2006**, *25*, 118−121. (c) Brandhorst, K.; Grunenberg, J. Efficient computation of compliance matrices in redundant internal coordinates from Cartesian Hessians for nonstationary points. *J. Chem. Phys.* **2010**, *132*, 184101−184107.

(29) Shahbazian, S. Why Bond Critical Points Are Not "Bond" Critical Points. *Chem. - Eur. J.* **2018**, *24*, 5401−5405.

(30) Weinhold, F.; Landis, C. R. *Valency and Bonding: A Natural Bond Orbital Donor - Acceptor Perspective*; Cambridge University Press: New York, 2005.

(31) Grimme, S.; Ehrlich, S.; Goerigk, L. Effect of the Damping Function in Dispersion Corrected Density Functional Theory. *J. Comput. Chem.* **2011**, *32*, 1456−1465.

(32) Bortolin, M.; Bucher, U. E.; Ruegger, H.; Venanzi, L. M.; Albinati, A.; Lianza, F.; Trofimenko, S. Synthesis and 2D NMR studies of {bis(pyrazolyl)borato} rhodium(I) complexes with weak Rh···H···C interactions and the x-ray crystal structure of {(cyclooctane-1, 5-diyl) bis (pyrazol-1-yl) borato}(1, 5-cyclooctadiene) rhodium. *Organometallics* **1992**, *11*, 2514−2521.

(33) Liu, X.; Pattacini, R.; Deglmann, P.; Braunstein, P. Do Short C-H-M (M= Cu (I), Ag (I)) Distances Represent Agostic Interactions in Pincer-Type Complexes? Unusual NHC Transmetalation from Cu (I) to Ag (I). *Organometallics* **2011**, *30*, 3302−3310.

4.4 *Publication:* Fast and Accurate Quantum Crystallography: from Small to Large, from Light to Heavy

The work with proteins in a quantum chemical context is computationally very demanding since wavefunctions can easily have several thousands of basis functions. While the calculation of ED might still be feasible in an overnight approach on a normal desktop computer, the evaluation of more complex functions like ELI or ESP will be very time-consuming. This is why the implementation of the calculation of the ESP was included in *cuQCrT*. Also, the evaluation of deformation density, giving insight into the description of asphericity in quantum crystallographic refinements, is a necessary output to look at the model quality. Therefore, the spherical densities of *Thakkar* [427, 428] were implemented into *cuQCrT* and the deformation density calculated in this publication. This showcases the scalability and possibilities of applications of *cuQCrT*, as this is the first time it was used on a complete protein. These implementations were used to calculate the ESP and deformation density and plot Figure 5 of this publication, which was completely created by me.

This publication shows how the fast calculation of wavefunctions enables applications for HAR, that were previously categorized as impossible or unfeasible for refinement. This refinement also shows how previous implementations of HAR struggled with the treatment of heavy elements. All this has been made easier in *NoSpherA2*, where the calculation of wavefunctions using GGA functionals instead of HF or hybrid functionals allows much faster wavefunction calculation and the description of relativistic basis sets and methods have been made available routinely. Additionally, the implementation of ELMO based refinements and the corresponding GUI are currently being tested, as the methods developed during this publication would be made more available through NoSpherA2.

Reprinted with permission from *Journal of Physical Chemistry Letters* **2019**, *10*, 22, 6973-6982. Copyright © 2019 American Chemical Society.

The Journal of Physical Chemistry Letters

Letter

Cite This: J. Phys. Chem. Lett. 2019, 10, 6973−6982

pubs.acs.org/JPCL

Fast and Accurate Quantum Crystallography: From Small to Large, from Light to Heavy

Lorraine A. Malaspina,[†] Erna K. Wieduwilt,[†,‡] Justin Bergmann,[†,○] Florian Kleemiss,[†,⊥] Benjamin Meyer,[‡,∥] Manuel F. Ruiz-López,[‡] Rumpa Pal,[†,#] Emanuel Hupf,[†,∇] Jens Beckmann,[†] Ross O. Piltz,[§] Alison J. Edwards,[§] Simon Grabowsky,[*,†,⊥] and Alessandro Genoni[*,‡]

[†] Institut für Anorganische Chemie und Kristallographie, Fachbereich 2 − Biologie/Chemie, Universität Bremen, Leobener Straße 3 und 7, 28359 Bremen, Germany

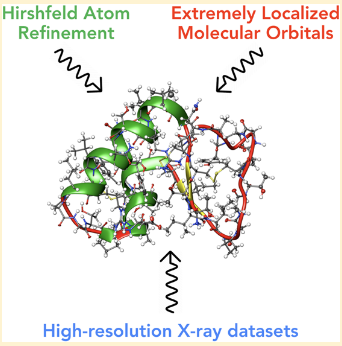
[‡] Université de Lorraine, CNRS, Laboratoire de Physique et Chimie Théoriques (LPCT), 1 Boulevard Arago, 57078 Metz, France

[⊥] Departement für Chemie und Biochemie, Universität Bern, Freiestrasse 3, 3012 Bern, Switzerland

[§] Australian Nuclear Science and Technology Organisation, Australian Centre for Neutron Scattering, New Illawarra Road, Lucas Heights, NSW 2234, Australia

Supporting Information

ABSTRACT: The coupling of the crystallographic refinement technique Hirshfeld atom refinement (HAR) with the recently constructed libraries of extremely localized molecular orbitals (ELMOs) gives rise to the new quantum-crystallographic method HAR-ELMO. This method is significantly faster than HAR but as accurate and precise, especially concerning the free refinement of hydrogen atoms from X-ray diffraction data, so that the first fully quantum-crystallographic refinement of a protein is presented here. However, the promise of HAR-ELMO exceeds large molecules and protein crystallography. In fact, it also renders possible electron-density investigations of heavy elements in small molecules and facilitates the detection and isolation of systematic errors from physical effects.



Structure determination via single-crystal X-ray diffraction is undoubtedly the most important technique for the experimental elucidation of the connectivity and three-dimensional structure of small as well as large molecules. Famous examples in the history of science are the X-ray structure determinations of biological molecules such as vitamins,[1] DNA,[2] or proteins.[3] The method is so successful and widespread that it gave rise to about 30 Nobel Prizes in chemistry, physics, and medicine.[3] Despite its paramount importance, it is sometimes forgotten that the refinement of the diffraction data is based on a crude approximation. In fact, the electron densities of the atoms are estimated as spheres (the so-called independent atom model, IAM[4]), which yields accurate atomic positions for all atoms except hydrogen but no direct information on chemical bonding and lone pairs, energies, or other properties of the molecules or materials under investigation. Moreover, if hydrogen atom positions are not attainable, intermolecular interactions such as hydrogen bonding or agostic interactions cannot be characterized adequately,[5] thermodynamic properties cannot be estimated,[6] and protonation states of enzymes remain unknown.[7] If the deformation of the atomic electron density due to chemical effects is neglected, bond properties in materials or in metalloproteins are in doubt.[8]

To solve such problems, on the experimental side, an enormous technical and financial effort has been made to obtain increasingly higher-quality and -resolution X-ray diffraction data sets, especially in protein crystallography. More and more dedicated end stations at synchrotron sources around the world provide a rapidly increasing number of high-resolution structures of proteins.[9,10] On the other hand, method development is trying to keep up by providing tools to replace the IAM with refinement techniques that take into account the asphericity of the atomic electron densities and that can potentially locate hydrogen atoms. For small molecules, multipole modeling has become a mature technique[11] and provides information on chemical bonding,[12] although refining hydrogen[13−15] and heavy atom[16,17] parameters is possible only in special cases. Unfortunately, for large molecules, multipole modeling is not feasible. However, fixed multipole parameters are stored in properly designed databanks[18−20] and successfully transferred to the protein structures of choice to perform accurate refinement of coordinates and atomic displacement parameters.[21,22]

Received: September 9, 2019
Accepted: October 21, 2019
Published: October 21, 2019

DOI: 10.1021/acs.jpclett.9b02646
J. Phys. Chem. Lett. 2019, 10, 6973−6982

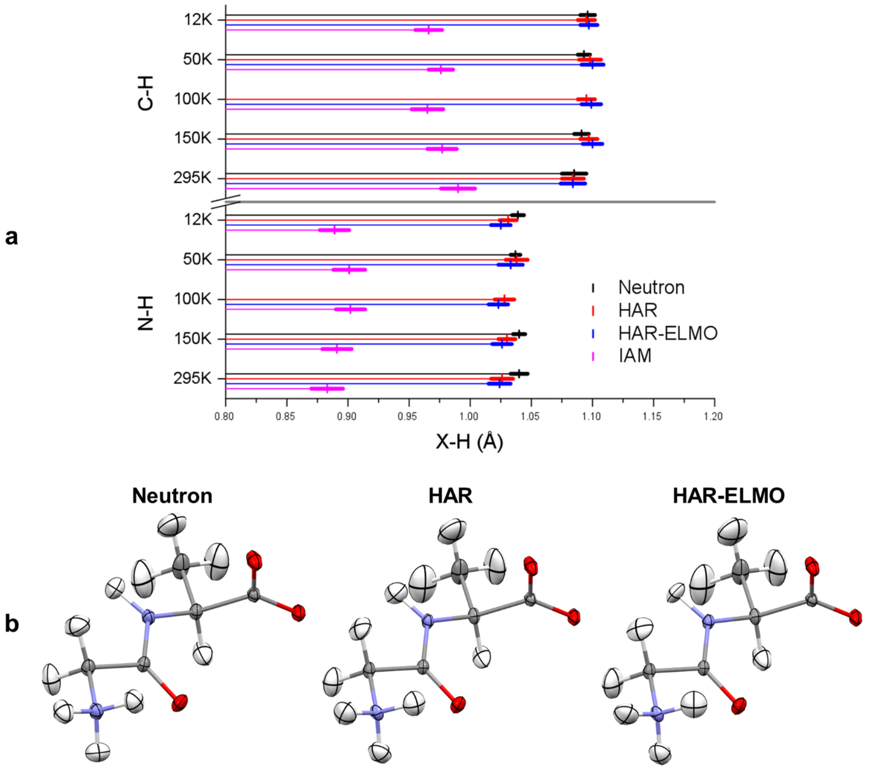


Figure 1. (a) Accuracy and precision of bond distances involving hydrogen atoms (X−H) for Gly-L-Ala at all temperatures (note that there is no neutron diffraction experiment at 100 K). Estimated accuracy is expressed by a match of the averaged bond lengths with the reference neutron experiment. Estimated precision is expressed by the size of the error bar, which represents the sample standard deviation upon averaging. (b) Visual comparison of refinement results of Gly-L-Ala at 150 K for neutron, HAR, and HAR-ELMO refinements. The remaining comparisons are depicted in the Supporting Information. Mean absolute differences (MADs) relative to the neutron experiment for Gly-L-Ala at 150 K: MAD($d_{\text{X−H,IAM}}$) = 0.128 Å, MAD($d_{\text{X−H,HAR-ELMO}}$) = 0.013 Å, MAD($d_{\text{X−H,HAR}}$) = 0.008 Å; MAD($U_{ij,\text{H,HAR-ELMO}}$) = 0.0079 Å^2, MAD($U_{ij,\text{H,HAR}}$) = 0.0078 Å^2.

Introducing quantum-chemical information into X-ray refinements to replace the IAM is an important aspect of quantum crystallography.[23,24] Within this framework, Hirshfeld atom refinement (HAR)[25,26] has been shown to be able to locate hydrogen atoms for small molecules from routine X-ray data with the same accuracy and precision as from neutron-diffraction experiments[27] and now is accessible in standard crystallographic refinement software.[28] Because it is based on regularly updated full wave function calculations, it is too slow to be applied to proteins or molecules containing heavy elements, the two compound classes that bear the most pressing problems of crystallography. For proteins, quantum-based refinements have been realized by utilizing restraints calculated quantum mechanically.[29−31]

In this work, we introduce a new method of quantum crystallographic refinement (HAR-ELMO) that combines HAR with the recently constructed libraries of extremely localized molecular orbitals (ELMOs) that rapidly provide wave functions of large and complicated molecules.[32] In fact, it has been shown previously that ELMOs for bonds and functional groups exhibit the same transferability from model compounds to proteins (or any other organic framework) as multipoles do, but with the significant advantage of allowing the reconstruction of full wave functions for the target systems.[33,34]

Here, we apply the new HAR-ELMO strategy to the amino acid L-alanine (L-Ala)[35] and the dipeptide glycyl-L-alanine (Gly-L-Ala)[26,36] to validate it against the original HAR technique and against suitable neutron-diffraction studies. Subsequently, we apply it to the polypeptide Leu-enkephalin,[37] the fibril-forming segment of the human prion protein,[38] and to crambin[39] to provide the first full and complete multipole-free quantum crystallographic refinement of a protein. This opens the door for a new research stream in protein crystallography. To demonstrate the potential applicability of HAR-ELMO in the field of coordination chemistry (e.g., organometallic compounds and metalloproteins), we apply it to the two coordination compounds $(Ph_3P)_2Hg(NO_3)_2$ and $(Ph_3P)_2Hg(CN)_2$, which are the first two fully quantum crystallographic refinements of molecules containing heavy elements.

Before applying the novel HAR-ELMO technique to proteins, we validated it on smaller systems, namely on Gly-L-Ala at five different temperatures (12, 50, 100, 150, and 295 K) and on L-Ala at three different temperatures (23, 100, and 150 K). The results clearly indicate that approximations introduced by the ELMO approach do not reduce the accuracy and precision of the HAR results (see Figure 1 and the Supporting Information). In fact, unlike IAM, C−H and N−H bond distances obtained at the HAR-ELMO level are as accurate (average values agree within one sample standard deviation (ssd) as shown in Figure 1a) and as precise (ssd values depicted as error bars in Figure 1a approximately agree) as the HAR and neutron results for all temperatures. This also holds for intermolecular hydrogen-bond parameters (Tables S14 and S15 in the Supporting Information). This means that

Table 1. Wall-Clock Timing on a Single CPU and Refinement Statistics[a]

structure	method	wall time (dd:hh:mm:ss)	$R_1(F)$	χ^2
Gly-L-Ala (150 K)	HAR	00:00:12:57	0.0161	2.336
	HAR-ELMO	00:00:03:36	0.0168	2.725
			(0.0251)	
L-Ala (150 K)	HAR	00:00:05:55	0.0202	3.152
	HAR-ELMO	00:00:02:25	0.0210	3.523
			(0.0279)	
Leu-enkephalin	HAR	00:09:52:00	0.0422	0.505
	HAR-ELMO	00:01:44:17	0.0430	0.545
			(0.0557)	
fibril-forming segment	HAR	01:07:00:00	0.0436	9.903
	HAR-ELMO	00:00:22:56	0.0474	11.913
			(0.0446)	
crambin (d = 0.54 Å)	HAR	impossible		
	HAR-ELMO	09:23:47:53	0.0715	5.004
			(0.0704)	
crambin (d = 0.73 Å)	HAR	impossible		
	HAR-ELMO	06:00:15:16	0.0624	7.672
			(0.0618)	
$(Ph_3P)_2Hg(NO_3)_2$	HAR	00:11:00:00	0.0162	4.232
	HAR-ELMO	00:00:31:47	0.0162	4.400
			(0.0204)	
$(Ph_3P)_2Hg(CN)_2$	HAR	00:13:30:00	0.0187	2.603
	HAR-ELMO	00:01:09:00	0.0187	2.622
			(0.0215)	

[a] The values in parentheses refer to the IAM R-values obtained after the refinement with the software TONTO. For Gly-L-Ala and L-Ala, only the 150 K measurements are given as examples (complete details are given in Table S11 of the Supporting Information).

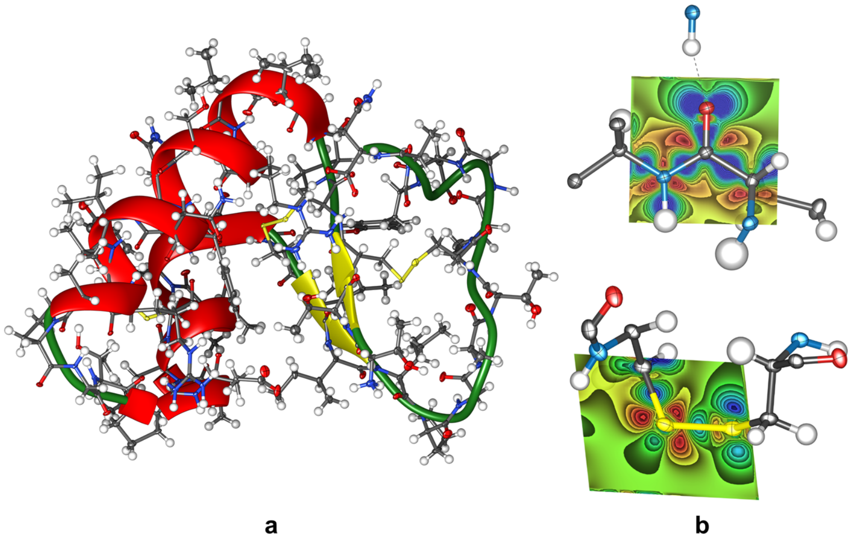


Figure 2. (a) Refined protein structure of crambin at d = 0.54 Å, not showing the disordered regions. For clarity, all H atoms are drawn with fixed spheres of 0.3 Å radius. Different representations for crambin at both selected resolutions and for the polypeptides are given in the Supporting Information. (b) Deformation density maps of crambin in a peptide region (C=O in Leu25) and a disulfide bond (between Cys4 and Cys32). Contour interval: 0.05 e$Å^{-3}$; blue, positive; red, negative; green, zero. For panels a and b, the ellipsoids are drawn at 50% probability.

the hydrogen atoms could be located successfully, but using a significantly lower wall-clock time on a single CPU (from now on, simply indicated as wall time) compared to traditional HAR (see Table 1).

HAR-ELMO even allows obtaining hydrogen anisotropic displacement parameters (ADPs) with precision and accuracy similar to those observed for the bond lengths. ADPs for hydrogen atoms are known to be important for the experimental determination of thermodynamic properties.[6] From visual inspection of Figure 1b, there is a qualitative agreement between the hydrogen and non-hydrogen ADPs obtained through neutron, HAR, and HAR-ELMO refinements of Gly-L-Ala at 150 K. This is confirmed quantitatively by extensive statistical analyses detailed in the Supporting Information for both compounds at all investigated temperatures.

DOI: 10.1021/acs.jpclett.9b02646
J. Phys. Chem. Lett. 2019, 10, 6973−6982

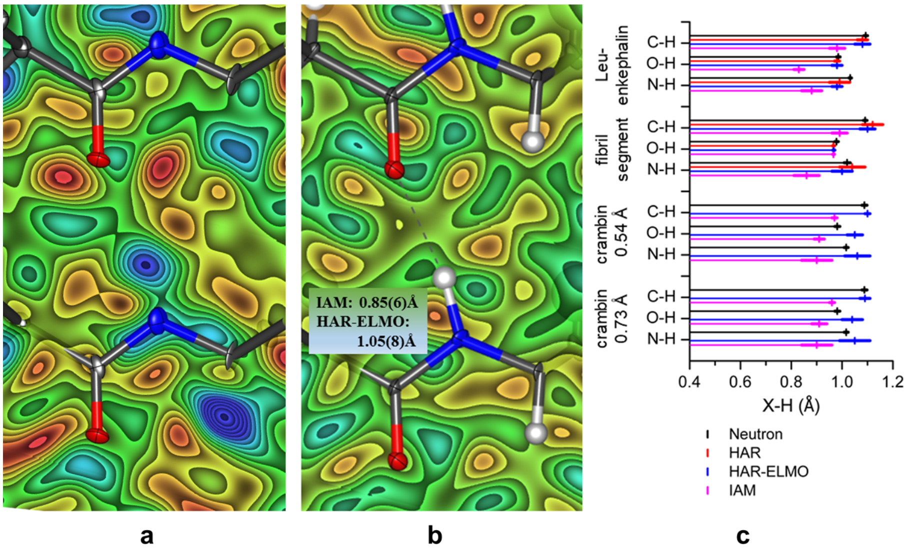


Figure 3. Residual electron density of the investigated fibril-forming segment (original PDB structure: 2OL9) in a region of intermolecular interaction between two parallel strands (namely, hydrogen-bond N2−H2···O2=C3 [according to the deposited CIF (Crystallographic Information File) labels] between residues Asn2 and Ser1): (a) IAM without hydrogen atoms as deposited in the original PDB and (b) HAR-ELMO with refined hydrogen atoms, including the IAM and HAR-ELMO refined N−H distances. For panels a and b, contour interval: 0.05 e$Å^{-3}$; blue, positive; red, negative; green, zero. For panels a and b, the ellipsoids are drawn at 50% probability. (c) Average values and associated standard deviations of C−H, O−H, and N−H bond lengths in IAM, HAR (if possible), and HAR-ELMO structures of the fibril-forming segment, Leu-enkephalin, and crambin compared to average values from neutron diffraction.[40]

After validation against amino acids and dipeptides, we applied HAR-ELMO to two polypeptides (Leu-enkephalin and the fibril-forming segment of the human prion protein) as well as to the protein crambin at two different resolutions (d = 0.54 and 0.73 Å). All the refinements successfully converged, yielding atomic coordinates and atomic displacement parameters for hydrogen atoms. Only in crambin it was required to constrain some atomic parameters mainly belonging to the disordered outermost and flexible loops of the protein (see Computational and Experimental Methods and the Supporting Information for more details). In Figure 2a the HAR-ELMO refined crambin structure is depicted, constituting the first quantum crystallographic structure of a protein. In Figure 2b, we show deformation density maps for a peptide bond fragment and a disulfide bond of crambin, which highlights the capability of the new HAR-ELMO method to describe bonding and lone-pair valence electron density distributions accurately. Residual and deformation densities for other regions of crambin and of the other two polypeptides can be found in the Supporting Information.

As a consequence of the improved modeling of bonding effects, the novel HAR-ELMO method is able to determine the positions of hydrogen atoms precisely and accurately, also in situations where they could not be detected previously. This can be seen in Figure 3, where, for the fibril-forming segment of the human prion protein, we show the model of a hydrogen bond interaction in the deposited PDB file (2OL9, Figure 3a) and upon HAR-ELMO refinement (Figure 3b). In the deposited model, all the hydrogen atoms are missing, although residual densities clearly indicate their presence (and the presence of related hydrogen bonds), whereas, after detecting them, we could refine them freely by exploiting the IAM and HAR-ELMO models. Modeled by IAM, the indicated N−H bond length is too short, whereas modeled by HAR-ELMO, it matches the reference neutron bond length of 1.03 Å.[40] Determination of accurate X-H bond lengths allowed us to calculate intermolecular interaction energies[41] between the fibril strands, which are a key to the investigation of the biological relevance of fibrils.[38] Parallel fibril strands are held together with a total interaction energy of 169 kJ/mol, whereby the electrostatic component and the dispersion component are about equal in size. This means that hydrogen bonding does not dominate the interaction inside the fibril. Overall, average bond lengths involving hydrogen atoms in the two polypeptides and crambin are significantly elongated with respect to the IAM refinements, and for all C−H bonds and most N−H and O−H bonds they agree with bond lengths derived from neutron diffraction[40] within a single sample standard deviation (see Figure 3c). For O−H and N−H bonds in crambin, the agreement is more problematic, which is independent of the resolution, but only depending on the data quality of protein crystallographic measurements (compare with Figure 1a for small-molecule crystallography).

The success of the new HAR-ELMO technique with respect to hydrogen atom localization derives from the fact that hydrogen has only one electron, which is always involved in bonding interactions, and it cannot be properly described with spherical atomic electron densities, as usually implemented in IAM. Overall, accounting for valence bonding deformations ensures a better agreement of the model with the measured diffraction pattern as long as the quality of the X-ray measurements includes this information. Table 1 shows, in terms of crystallographic R values, that this is the case for Gly-L-Ala, L-Ala, Leu-enkephalin, and the mercury complexes that will be discussed below. For Leu-enkephalin, the improvement is significant (from R_{IAM} = 0.0557 to $R_{\mathrm{HAR\text{-}ELMO}}$ = 0.0430), and Figure S87 shows that this is due to residual electron density in the bonding and lone-pair regions after IAM. For crambin and the fibril-forming segment of the human prion protein, the quality of the measurements is not sufficient to observe valence electron density (see Figures 3a and 3b), so there is no improvement in *R*-values, but only in refined structural parameters and in electronic properties contained in the wave function.

The more sensitive χ^2 values in Table 1 show that HAR is always in slightly better agreement with the measured diffraction pattern than HAR-ELMO because of the approximations used in the ELMO procedure, but this difference does not reduce the accuracy of geometrical and derived properties as shown above. This is important because, for crambin, and consequently every larger molecule, HAR is impossible or impractical because it is too computationally demanding, requiring the recalculation of the wave function after every refinement step (see Table 1). Therefore, HAR-ELMO renders quantum crystallographic refinement possible for protein crystallography. The computational cost depends on the size of the molecule (total number of atomic orbitals used in the calculation) and on the number of structure factors used in the least-squares refinement. For crambin, the number of reflections in the higher-resolution data set (d = 0.54 Å, 96139 refl.) is more than twice as high as the number in the lower-resolution data set (d = 0.73 Å, 45265 refl.). The HAR-ELMO wall time reduces from almost 10 days to about 6 days (see Table 1) from the higher- to the lower-resolution data set.

The reduction of computational cost from HAR to HAR-ELMO for the two coordination compounds $(Ph_3P)_2Hg(NO_3)_2$ and $(Ph_3P)_2Hg(CN)_2$ is significant (22- and 13-fold reduction, respectively; Table 1). Figure 4c also shows that the hydrogen atom positions and ADPs in the phenyl groups of $(Ph_3P)_2Hg(NO_3)_2$ are accurately determined, even with the heavy mercury dominating the scattering. A comparison of refinement results with isotropic hydrogen atom displacement parameters and with $(Ph_3P)_2Hg(CN)_2$ is given in the Supporting Information.

In the IAM refinement (Figure 4b), there is a significant amount of unmodelled physically meaningful residual density in the C−C, C−P, N−O, Hg−O, and Hg−P bonds as well as in the P and Hg core regions. The maximum unmodelled electron density peak in the Hg core is +2.1 eÅ^{-3}. The HAR-ELMO refinement accounts for the deformation of the electron density caused by bonding, so unmodelled bonding features are absent from the residual-density map after HAR-ELMO (Figure 4c). The maximum unmodelled core Hg density reduces to +1.6 eÅ^{-3}, a significant reduction showing the importance of modeling the aspherical core deformation of heavy mercury. Similar, but less pronounced, features can be seen for $(Ph_3P)_2Hg(CN)_2$, described in the Supporting Information (maximum residual density for this compound being +1.3 eÅ^{-3} (IAM) and +1.1 eÅ^{-3} (HAR-ELMO)). HAR gives very similar results in terms of residual electron density (+1.7 and +1.1 eÅ^{-3} for $(Ph_3P)_2Hg(NO_3)_2$ and $(Ph_3P)_2Hg(CN)_2$, respectively).

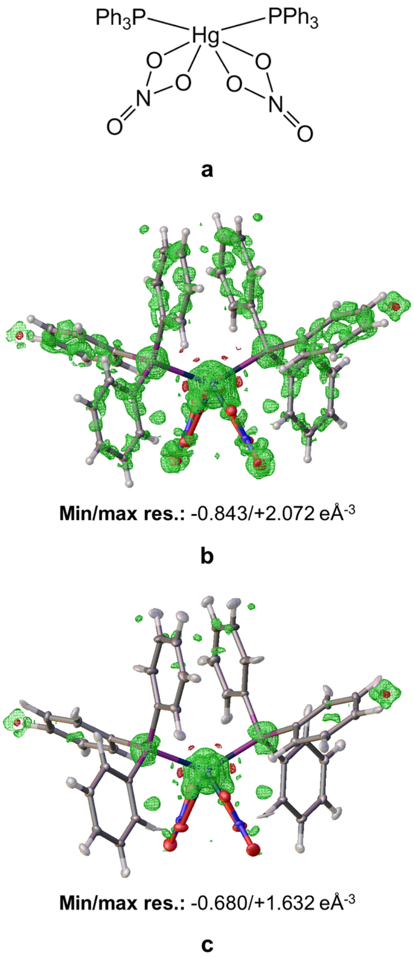


Figure 4. Refined structures of (a) $(Ph_3P)_2Hg(NO_3)_2$ with residual electron density distribution and minimum/maximum value (b) after IAM and (c) after HAR-ELMO. Isovalue = ± 0.2 eÅ^{-3}; green, positive; red, negative. ADPs at 50% probability level. The two nitrate and triphenylphosphine groups are symmetry-equivalent with their counterpart leading to 6-fold coordination of Hg. Distances after HAR-ELMO: Hg−P = 2.4502(5) Å, Hg−O = 2.4692(7) Å, and Hg−O′ = 2.8446(9) Å.

The residual density distribution after HAR-ELMO shown in Figure 4c is not free of systematic effects. Unmodelled effects remain at the Hg and P cores, and there are two ghost peaks that are not due to aspherical bonding deformation. The HAR-ELMO treatment allows examination of these additional effects. The ghost peaks may be Fourier truncation errors that we could not remove in the data reduction procedure. The features at the Hg core are physical effects that the HAR-ELMO model (including anomalous dispersion and absorption corrections) did not resolve. The most obvious problems in our theoretical ansatz are the neglect of electron correlation and relativistic effects. Therefore, we performed fully relativistic HARs at the B3LYP-IOTC level,[42] which is currently not possible for HAR-ELMO. This refinement took 8 days and 19 hours and only reduced the positive maximum residual density to 1.15 eÅ^{-3}, still leaving very similar systematic features at the Hg core (see Figure S97). Other possible effects could be anharmonicity, insufficiently modeled absorption, or radiation damage. All these effects can be investigated in detail in future projects because the fast HAR-ELMO refinements allowed us to separate deformation due to bonding from all the other important effects.

Considering all the results shown above, the HAR-ELMO structural model includes chemically meaningful bonding and nonbonding features, in contrast to IAM. As long as the quality of the diffraction experiment is high enough to detect such features, the new HAR-ELMO refinement method leads to significantly improved agreements between the model and the measured diffraction data, as reflected by reduced values of figures of merit (R values, χ^2, and residual density

DOI: 10.1021/acs.jpclett.9b02646
J. Phys. Chem. Lett. 2019, 10, 6973−6982

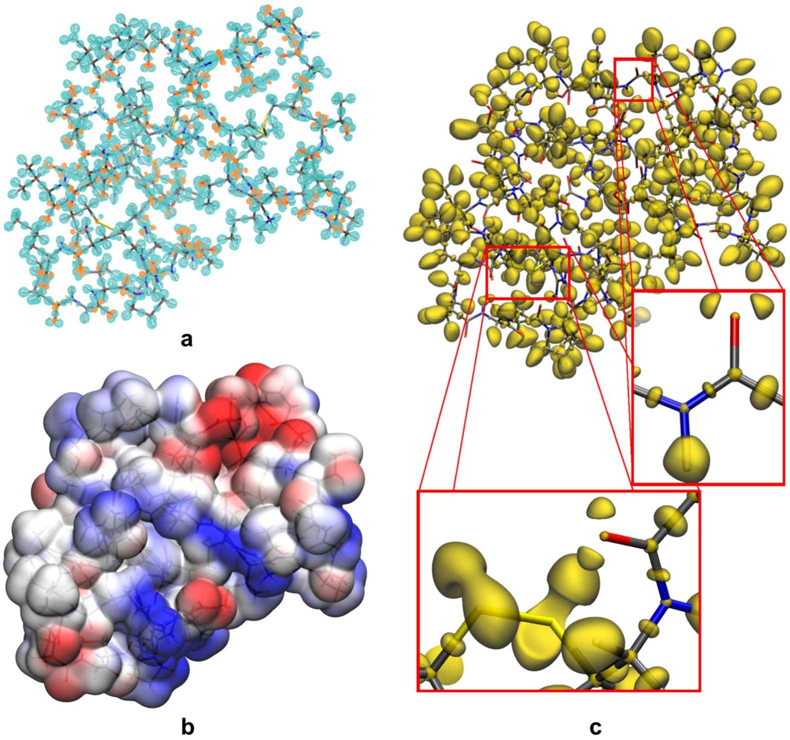


Figure 5. cuQCT GPU-accelerated calculations of molecular properties for crambin based on the final HAR-ELMO geometry and wave function: (a) isosurfaces of deformation density at ±0.04 eÅ^{-3} (blue, positive; red, negative); (b) electrostatic potential in a range from −0.1 to +0.1 e bohr^{-1} mapped on an electron-density isosurface of 0.001 e bohr^{-3}; (c) ELI-D at an isovalue of 2.2 for the main picture and 2.0 for the magnified details. All grids were calculated within 2 hours on a normal desktop computer with a grid-point separation of 0.1 bohr, except for the ESP grid which took about 500 hours at a 0.6 bohr separation. The program cuQCT is a software for the rapid calculation of large and dense grids of molecular properties on GPUs developed by some of us at the Universities of Bremen and Bern.

distributions). When the quality of the measurements is not high enough, the intrinsic quantum-mechanical nature of the underlying model nevertheless allows the derivation of accurate experimental geometric parameters even in cases where ADPs become nonpositive definite in free HAR and free HAR-ELMO refinements and can be used to obtain properties that are not restricted to those associated with the electron density of the system under consideration. Accessible properties are, for example, bonding descriptors such as the electron localizability indicator (ELI-D),[43] electrostatic potentials (ESPs), noncovalent interaction (NCI) plots,[44,45] or inter- and intramolecular interaction energies, all of which are important for rational drug design.[46−48] Figure 5 shows the deformation density, ESP, and the ELI-D for the entire crambin molecule, calculated rapidly on graphics cards using our own GPU-accelerated software cuQCT, based on the wave function and geometry from the last HAR-ELMO step.

For all the mentioned properties, accurate hydrogen atom positions are imperative. This study shows that HAR-ELMO can detect hydrogen atoms and locate them accurately and precisely, even for those large systems for which a fully quantum crystallographic refinement was previously considered prohibitive.

Notwithstanding the promise of the current results, the major drawback associated with protein refinements is solvent and intrinsic disorder. At the moment, there is no refinement engine underlying original quantum crystallography methods such as HAR that can treat this problem in a proper crystallographic way. It will be of utmost importance to develop and implement a novel strategy to refine disorder in HAR-ELMO or couple HAR-ELMO to existing software for protein refinement that already provides the solution. Although nowadays the number of exceptionally high-resolution protein structures is increasing, our study shows that the results of the HAR-ELMO strategy do not depend on using very high resolution data (0.85 Å is sufficient) but rather on the quality of the data.

To improve the current version of the HAR-ELMO technique, on the one hand, we have already envisaged the implementation of a parallelized version of the presented HAR-ELMO procedure, also introducing linear-scaling options for the computation of one-electron density matrices. This will result in a further and drastic reduction of the computational cost. On the other hand, we are also developing a multiscale HAR-QM/ELMO technique. It will consist in coupling HAR with the QM/ELMO method,[49] a recently introduced embedding strategy that describes the most chemically/biologically relevant part of the system (e.g., the active site of a protein or heavy elements in coordination compounds) at a very high QM level (e.g., including dispersion for weak intermolecular interactions in enzyme pockets, or core excitations and relativity for heavy elements), while treating the rest of the system with fixed ELMOs transferred from the databanks. This will increase the accuracy of the results for the regions of particular interest while keeping the computational cost within reasonable limits.

The HAR-ELMO refinement captures aspherical electron-density features, such as bonds, lone pairs and the core

DOI: 10.1021/acs.jpclett.9b02646
J. Phys. Chem. Lett. 2019, 10, 6973−6982

deformation of heavy elements, in a rapid and automated way. This allows the separation of such physical effects from unmodelled systematic effects, which are smaller but still significant. These unmodelled effects can be of two different natures: (i) effects that are related to the adopted model (e.g., relativity, basis-set dependency and anharmonicity) or (ii) effects related to the measurement and data reduction strategy. To tackle problem (i) we propose applying the above-mentioned HAR-QM/ELMO strategy. Furthermore, we can exploit the speed of HAR-ELMO and HAR-QM/ELMO in order to use them as debugging tools to address problem (ii). In this way, changes in data-reduction routines (for instance, absorption correction or mitigation of radiation damage) can directly be mapped on the electron density.

While traditional refinement strategies normally place hydrogen atoms according to valence considerations and not according to measured electron-density peaks and while the pioneering and innovative QM/MM refinements of proteins[29−31] are still based on the use of tabulated spherical atomic form factors, the new HAR-ELMO technique explicitly treats for the first time the asphericity of the atomic electron densities in a quantum mechanical way to compute structure factors for the refinement of macromolecular crystal structures. This adds unprecedented insights into protein structure determination. Moreover, for the chemical-crystallographic community, a major advantage of the new approach also lies in the future possibility of extending the ELMO libraries to very different classes of compounds, such as protein inhibitors or ligands that regularly occur in syntheses of, for example, metal and covalent organic frameworks (MOFs and COFs). In particular, if ligands are sterically demanding or contain heavy elements in a closed-shell state, resorting to such ELMO libraries will render quantum crystallographic refinements possible for high-throughput investigations of chemical bonding features.

In conclusion, two classes of compounds can now be modeled by quantum crystallography because of the developments presented here: proteins and molecules containing heavy elements. For large molecules, the treatment of the lightest atoms is crucial; for small molecules, the treatment of the heaviest elements is the biggest challenge. HAR-ELMO is today the only method that allows fully quantum-crystallographic refinements of both such very different compound types on a reasonable time frame. This reduction of computational cost derives from the introduction of the ELMO approximation in the procedure, but despite this, accuracy and precision are only marginally affected. In future work, we will exploit the new possibilities in macromolecular chemistry to investigate bonding situations in metalloproteins and for enzyme−ligand binding as well as in chemical crystallography to accurately model bonding to heavy elements and understand unusually close intermolecular contacts, e.g., in halogen and hydrogen bonding motifs.

■ COMPUTATIONAL AND EXPERIMENTAL METHODS

In this section, all the computational and experimental methods used in the present investigation are briefly presented. More details are given in the Supporting Information.

HAR-ELMO Implementation. The HAR-ELMO technique was implemented interfacing the HAR functionalities in the software TONTO[50] with the ELMOdb program associated with the ELMO libraries.[32] The interface between these two programs is the lamaGOET graphical user interface (GUI), which reads the initial geometry (PDB, CIF, or XYZ file) from a previous IAM refinement and provides it to ELMOdb for the automatic transfer of molecular orbitals. After the transfer, the wave function and first-order density matrix of the system are computed and written to a formatted checkpoint file, which is afterward passed to TONTO by lamaGOET together with the ADPs of the IAM refinement, in order to perform the first HAR iteration. The updated geometry and ADPs are used as input for iteratively repeating the procedure until convergence. Although interfaced, lamaGOET and ELMOdb are currently two independent programs. The former is available free of charge at http://www.tinyurl.com/lamaGOET (source code is also available on https://github.com/lomalaspina/lamaGOET). In contrast, at the moment, the ELMO libraries and the associated ELMOdb program are not publicly available, but they can be obtained upon motivated request to A.G. (Alessandro.Genoni@univ-lorraine.fr). The construction of a dedicated webserver is already envisaged.

ELMO Libraries. The used ELMO libraries[32] cover all the functional units of water and of the 20 natural amino acids in all their possible protonation states and forms. The associated ELMOdb program can also read tailor-made ELMOs properly computed to describe fragments of specific molecules or ligands beyond the standard amino acids and the water molecule. Other than selecting the molecular orbitals associated with the fragments of the system under examination, ELMOdb also automatically handles the rotation of the ELMOs by simply reading the geometry of the target molecule, without the need for defining any local coordinate systems as usually done in multipole refinements. In particular, following a method originally introduced by Philipp and Friesner,[51] this is done by the program defining, for each fragment, a rotation matrix from the reference frame in the model molecule (on which the ELMOs were originally computed) to the reference frame in the target system (see the Supporting Information for more details).

Additional Refinement Details. For the HAR-ELMO refinements of systems that involve only amino acids and water molecules, we exploited the ELMO library for the 6-311G(d,p) basis set. For the HAR-ELMO refinements of the two coordination compounds we used tailor-made ELMOs computed on proper model molecules with the DZP-DKH basis-set. For comparison, we performed traditional HARs at the respective same basis sets and at the Hartree−Fock level, except for crambin for which traditional HAR is too time-consuming. In addition, for $(PPh_3)_2Hg(NO_3)_2$ a HAR at fully relativistic infinite-order two-component level (IOTC) was performed with the DZP-DKH basis-set.[42] All refinements were performed in TONTO 18.12.17 github v. 2b3dd9c.

Validation of HAR-ELMO. The experimental X-ray[26] and neutron[36] data of Gly-L-Ala were obtained from the literature. For L-Ala, the X-ray data at 23 K were taken from Destro et al.,[35] whereas the data at 100 and 150 K were measured using a Bruker Venture D8 diffractometer with a Mo Kα microsource and a Photon100 CMOS detector to a maximum resolution of 0.45 Å (100 K) and 0.55 Å (150 K). Neutron-diffraction data of L-Ala at 23 K and 150 K were measured at the Australian Centre for Neutron Scattering of ANSTO (beamline KOALA) using the Laue diffraction technique.[52] KOALA is equipped with a large cylindrical image plate, so that a resolution of 0.50 Å was reached in the wavelength range 0.750 to 1.500 Å, refined to R values $R(F)$= 0.033/0.038. In addition to the

Supporting Information, more details are in the Cambridge Structural Database (CSD) under nos. CCDC 1917810–1917812, 1917816–1917818, 1917820–1917822, and 1917823–1917837 (X-ray) and CCDC 1917685 and 1917813 (neutron).

Polypeptide and Protein Refinements. The experimental data of crambin were taken from the literature.[39] In the available PDB file, there was an unspecified and unresolved amount of disordered water molecules. Because disorder for HAR, HAR-ELMO, or any other refinement belonging to the original definition of quantum crystallography is currently an unsolved problem, we used the SQUEEZE routine[53] to calculate the scattering contributions of the solvent molecules and excluded them from the original structure factors. These structure factors were used for full-matrix least-squares IAM and HAR-ELMO refinements in TONTO at two resolutions, 0.54 and 0.73 Å. Parameters of atoms belonging to lower-occupancy regions in the reported crambin structure were removed, and only the major component atoms were included at full occupancy. Positional and displacement parameters were freely refined for all atoms, except those in the disorder regions and those of some atoms belonging to the outermost flexible loop of crambin. For those atoms, positions and displacement parameters were kept fixed at values obtained from an initial unconstrained IAM refinement.

The original structure in the 2OL9 PDB file for the fibril-forming segment of the human prion protein is resolved up to 0.85 Å and contains ordered oxygen sites of water solvent molecules, but no hydrogen atoms.[38] Before the initial IAM refinement, all hydrogen atoms were geometrically placed and then refined freely in the subsequent HAR and HAR-ELMO refinements. The experimental data of Leu-enkephalin[37] are resolved to $d = 0.43$ Å, and all atoms are present and ordered in the original file. The reference IAM refinement was performed in the software TONTO and followed by both HAR and HAR-ELMO treatments. The files associated with the polypeptide refinements correspond to CSD depositions CCDC 1917596–1917598 and 1917730–1917732.

Small-Molecule Refinement Involving Heavy Elements. The two coordination compounds $(Ph_3P)_2Hg(NO_3)_2$ and $(Ph_3P)_2Hg(CN)_2$ are known from the literature,[54] but they were resynthesized according to the literature procedures, crystallized, and measured at the same X-ray setup described above to resolutions of 0.50 and 0.55 Å, respectively. $(Ph_3P)_2Hg(NO_3)_2$ is the only compound in this study with half a molecule in the asymmetric unit ($Z' = 0.5$), whereas for all the others $Z' = 1$. IAM, HAR, and HAR-ELMO refinements were performed in TONTO. For HAR and HAR-ELMO, two models were adopted: one with anisotropic and one with isotropic hydrogen displacement parameters. CIFs of the isotropic HARs were deposited with the CSD under deposition numbers CCDC-1917593 and 1917594. CIFs and structure factor files for all models are available from the Supporting Information of this Letter.

■ ASSOCIATED CONTENT

Ⓢ Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jpclett.9b02646.

Theoretical bases of the HAR-ELMO technique, details about the validation of the HAR-ELMO method, details and further results of the polypeptides and protein refinements, and details and further results of the refinements of the coordination compounds (PDF)

Crystallographic information files (.cif files) and files containing experimental and calculated structure factors (.fco files) associated with all the refinements (all the considered models) for all the compounds in this study (ZIP)

■ AUTHOR INFORMATION

Corresponding Authors

*Phone: +33 (0)3 72 74 91 70. Fax: +33 (0)3 72 74 91 87. E-mail: Alessandro.Genoni@univ-lorraine.fr.
*Phone: +41 31 631 4281. E-mail: simon.grabowsky@dcb.unibe.ch.

ORCID

Manuel F. Ruiz-López: 0000-0002-2784-6319
Rumpa Pal: 0000-0003-3080-3333
Emanuel Hupf: 0000-0003-0121-7554
Jens Beckmann: 0000-0002-8548-1821
Ross O. Piltz: 0000-0003-0867-3712
Simon Grabowsky: 0000-0002-3377-9474
Alessandro Genoni: 0000-0002-3511-4855

Present Addresses

○J. Bergmann: Lund University, Department of Theoretical Chemistry, Chemical Center, P.O. Box 124, S-22100 Lund, Sweden.

∥B. Meyer: Laboratory for Computational Molecular Design (LCMD), Institute of Chemical Sciences and Engineering (ISIC), École Polytechnique Fédérale de Lausanne (EPFL), CH-1015 Lausanne, Switzerland.

#R. Pal: Division of Physics, Faculty of Pure and Applied Sciences, University of Tsukuba, 1-1-1 Tennodai, Tsukuba, Ibaraki 305-8571, Japan.

∇E. Hupf: Department of Chemistry, University of Alberta, 11227 Saskatchewan Drive, Edmonton, Alberta, Canada T6G 2G2.

Author Contributions

A.G. and S.G. conceived the HAR-ELMO method and secured funding for the project, the latter together with J. Beckmann. L.A.M., A.G., and S.G. designed the study, and A.G. and S.G. supervised it and wrote the Letter. L.A.M. developed the software lamaGOET to couple HAR to the ELMO libraries with the help of A.G. L.A.M. performed all the structural refinements, obtained the results for this Letter, and generated most of the display items. L.A.M. and E.K.W. performed the statistical analyses. E.K.W. collated the Supporting Information file and generated Figure 3c. E.H. synthesized and crystallized both Hg-containing compounds. R.P. measured the data set of $(Ph_3P)_2Hg(NO_3)_2$. J. Bergmann and L.A.M. measured the data set of $(Ph_3P)_2Hg(CN)_2$ and performed the relativistic refinement of $(Ph_3P)_2Hg(NO_3)_2$ together. F.K. developed the code to calculate electrostatic potential, deformation density, and electron localizability in cuQCT, carried out the cuQCT calculations, and generated Figure 5. J. Beckmann and M.F.R.-L. helped to design the study and interpret the results. B.M. helped to generate some material for the Supporting Information. E.K.W., L.A.M., and R.P. measured the 23 K neutron-diffraction data set of L-alanine under the supervision of R.O.P. The 150 K neutron-diffraction data set of L-alanine was measured by A.J.E. The 100 and 150 K X-ray data sets of L-alanine were measured by L.A.M.

Notes

The authors declare no competing financial interest.

Data Availability: The best-quality models for all compounds obtained in the present study (except those for crambin) have been deposited at the Cambridge Crystallographic Data Centre (CCDC) under the deposition numbers indicated in the Computational and Experimental Methods section. This data can be obtained free of charge via www.ccdc.cam.ac.uk/data_request/cif. All models for all refinements (including the inferior models and including the protein crambin) plus the measured structure factors have been deposited with this journal in addition to the CCDC deposition to ensure reproducibility of the study. The lamaGOET GUI can be obtained free of charge for registered users at http://www.tinyurl.com/lamaGOET. All other programs developed for the present study are available from the corresponding authors of the paper upon motivated request.

■ ACKNOWLEDGMENTS

S.G. thanks the German Research Foundation (Deutsche Forschungsgemeinschaft DFG) for funding within the projects GR 4451/1-1 and GR 4451/2-1, J.B. within BE 3716/7-1, and E.H. for a research fellowship HU 2512/1-1. A.G. acknowledges the French Research Agency (ANR) for financial support of the Young Researcher Project *QuMacroRef* through Grant No. ANR-17-CE29-0005-01. The neutron-diffraction experiments of L-alanine were performed on the KOALA beamline of the OPAL reactor under the Australian Centre for Neutron Scattering proposal no. 6049. The authors acknowledge Dylan Jayatilaka (University of Western Australia) for helpful discussions, Benoît Guillot (University of Lorraine) for providing experimental data of the Leu-enkephalin polypeptide, and Lukas Bučinský (Slovak University of Technology) for providing assistance with the relativistic refinement.

■ REFERENCES

(1) Hodgkin, D. C.; Kamper, J.; Mackay, M.; Pickworth, J.; Trueblood, K. N.; White, J. G. Structure of Vitamin B_{12}. *Nature* **1956**, *178*, 64−66.

(2) Watson, J. D.; Crick, F. H. C. Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid. *Nature* **1953**, *171*, 737−738.

(3) Jaskolski, M.; Dauter, Z.; Wlodawer, A. A brief history of macromolecular crystallography, illustrated by a family tree and its Nobel fruits. *FEBS J.* **2014**, *281*, 3985−4009.

(4) Compton, A. H. The Distribution of Electrons in Atoms. *Nature* **1915**, *95*, 343−344.

(5) Braga, D.; Grepioni, F.; Biradha, K.; Desiraju, G. R. Agostic interactions in organometallic compounds. A Cambridge Structural Database study. *J. Chem. Soc., Dalton Trans.* **1996**, 3925−3930.

(6) Hoser, A. A.; Madsen, A. Ø. Dynamic quantum crystallography: lattice dynamical models refined against diffraction data. II. Applications to *L*-alanine, naphthalene and xylitol. *Acta Crystallogr., Sect. A: Found. Adv.* **2017**, *73*, 102−114.

(7) Fisher, S. J.; Blakeley, M. P.; Cianci, M.; McSweeney, S.; Helliwell, J. R. Protonation-state determination in proteins using high-resolution X-ray crystallography: effects of resolution and completeness. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2012**, *68*, 800−809.

(8) Lu, Y.; Yeung, N.; Sieracki, N.; Marshall, N. M. Design of functional metalloproteins. *Nature* **2009**, *460*, 855−862.

(9) Petrova, T.; Podjarny, A. Protein crystallography at subatomic resolution. *Rep. Prog. Phys.* **2004**, *67*, 1565−1605.

(10) Ogata, H.; Nishikawa, K.; Lubitz, W. Hydrogens detected by subatomic resolution protein crystallography in a [NiFe] hydrogenase. *Nature* **2015**, *520*, 571−574.

(11) Coppens, P. Charge Densities Come of Age. *Angew. Chem., Int. Ed.* **2005**, *44*, 6810−6811.

(12) Stalke, D. *Electron Density and Chemical Bonding I. Experimental Charge Density Studies*; Springer-Verlag: Berlin & Heidelberg, 2012.

(13) Hoser, A. A.; Dominiak, P. M.; Woźniak, K. Towards the best model for H atoms in experimental charge-density refinement. *Acta Crystallogr., Sect. A: Found. Crystallogr.* **2009**, *65*, 300−311.

(14) Zhurov, V. V.; Zhurova, E. A.; Stash, A. I.; Pinkerton, A. A. Importance of the consideration of anharmonic motion in charge-density studies: a comparison of variable-temperature studies on two explosives, RDX and HMX. *Acta Crystallogr., Sect. A: Found. Crystallogr.* **2011**, *67*, 160−173.

(15) Dittrich, B.; Lübben, J.; Mebs, M.; Wagner, A.; Luger, P.; Flaig, R. Accurate Bond Lengths to Hydrogen Atoms from Single-Crystal X-ray Diffraction by Including Estimated Hydrogen ADPs and Comparison to Neutron and QM/MM Benchmarks. *Chem. - Eur. J.* **2017**, *23*, 4605−4614.

(16) Iversen, B. B.; Larsen, F. K.; Pinkerton, A. A.; Martin, A.; Darovsky, A.; Reynolds, P. A. Accurate charge densities in days − use of synchrotrons, image plates and very low temperatures. *Acta Crystallogr., Sect. B: Struct. Sci.* **1999**, *55*, 363−374.

(17) Zhurov, V. V.; Zhurova, E. A.; Stash, A. I.; Pinkerton, A. A. Characterization of Bonding in Cesium Uranyl Chloride: Topological Analysis of the Experimental Charge Density. *J. Phys. Chem. A* **2011**, *115*, 13016−13023.

(18) Domagała, S.; Fournier, B.; Liebschner, D.; Guillot, B.; Jelsch, C. An Improved Experimental Databank of Transferable Multipolar Atom Models − ELMAM2. Construction Details and Applications. *Acta Crystallogr., Sect. A: Found. Crystallogr.* **2012**, *68*, 337−351.

(19) Dominiak, P. M.; Volkov, A.; Li, X.; Messerschmidt, M.; Coppens, P. A Theoretical Databank of Transferable Aspherical Atoms and Its Application to Electrostatic Interaction Energy Calculations of Macromolecules. *J. Chem. Theory Comput.* **2007**, *3*, 232−247.

(20) Dittrich, B.; Hübschle, C. B.; Pröpper, K.; Dietrich, F.; Stolper, T.; Holstein, J. J. The Generalized Invariom Database (GID). *Acta Crystallogr., Sect. B: Struct. Sci., Cryst. Eng. Mater.* **2013**, *69*, 91−104.

(21) Elias, M.; Liebschner, D.; Koepke, J.; Lecomte, C.; Guillot, B.; Jelsch, C.; Chabriere, E. Hydrogen atoms in protein structures: high-resolution X-ray diffraction structure of the DFPase. *BMC Res. Notes* **2013**, *6*, 308.

(22) Pröpper, K.; Holstein, J. J.; Hübschle, C. B.; Bond, C. S.; Dittrich, B. Invariom refinement of a new monoclinic solvate of thiostrepton at 0.64 Å resolution. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2013**, *69*, 1530−1539.

(23) Grabowsky, S.; Genoni, A.; Bürgi, H.-B. Quantum Crystallography. *Chem. Sci.* **2017**, *8*, 4159−4176.

(24) Genoni, A.; Bučinský, L.; Claiser, N.; Contreras-García, J.; Dittrich, B.; Dominiak, P. M.; Espinosa, E.; Gatti, C.; Giannozzi, P.; Gillet, J.-M.; Jayatilaka, D.; Macchi, P.; Madsen, A. Ø.; Massa, L. J.; Matta, C. F.; Merz, K. M., Jr.; Nakashima, P. N. H.; Ott, H.; Ryde, U.; Schwarz, K.; Sierka, M.; Grabowsky, S. Quantum Crystallography: Current Developments and Future Perspectives. *Chem. - Eur. J.* **2018**, *24*, 10881−10905.

(25) Jayatilaka, D.; Dittrich, B. X-ray structure refinement using aspherical atomic density functions obtained from quantum-mechanical calculations. *Acta Crystallogr., Sect. A: Found. Crystallogr.* **2008**, *64*, 383−393.

(26) Capelli, S.; Bürgi, H.-B.; Dittrich, B.; Grabowsky, S.; Jayatilaka, D. Hirshfeld atom refinement. *IUCrJ* **2014**, *1*, 361−379.

(27) Woińska, M.; Grabowsky, S.; Dominiak, P. M.; Woźniak, K.; Jayatilaka, D. Hydrogen atoms can be located accurately and precisely by X-ray crystallography. *Sci. Adv.* **2016**, *2*, e1600192.

(28) Fugel, M.; Jayatilaka, D.; Hupf, E.; Overgaard, J.; Hathwar, V. R.; Macchi, P.; Turner, M. J.; Howard, J. A. K.; Dolomanov, O. V.; Puschmann, H.; Iversen, B. B.; Bürgi, H.-B.; Grabowsky, S. Probing

DOI: 10.1021/acs.jpclett.9b02646
J. Phys. Chem. Lett. 2019, 10, 6973−6982

the accuracy and precision of Hirshfeld atom refinement with HARt interfaced with Olex2. *IUCrJ* **2018**, *5*, 32−44.

(29) Zheng, M.; Reimers, J. R.; Waller, M. P.; Afonine, P. V. Q|R: quantum-based refinement. *Acta Cryst. D* **2017**, *73*, 45−52.

(30) Ryde, U.; Olsen, L.; Nilsson, K. Quantum chemical geometry optimizations in proteins using crystallographic raw data. *J. Comput. Chem.* **2002**, *23*, 1058−1070.

(31) Yu, N.; Yennawar, H. P.; Merz, K. M., Jr. Refinement of protein crystal structures using energy restraints derived from linear scaling quantum mechanics. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2005**, *61*, 322−332.

(32) Meyer, B.; Genoni, A. Libraries of Extremely Localized Molecular Orbitals. 3. Construction and Preliminary Assessment of the New Databanks. *J. Phys. Chem. A* **2018**, *122*, 8965−8981.

(33) Meyer, B.; Guillot, B.; Ruiz-Lopez, M. F.; Genoni, A. Libraries of Extremely Localized Molecular Orbitals. 1. Model Molecules Approximation and Molecular Orbitals Transferability. *J. Chem. Theory Comput.* **2016**, *12*, 1052−1067.

(34) Meyer, B.; Guillot, B.; Ruiz-Lopez, M. F.; Jelsch, C.; Genoni, A. Libraries of Extremely Localized Molecular Orbitals. 2. Comparison with the Pseudoatoms Transferability. *J. Chem. Theory Comput.* **2016**, *12*, 1068−1081.

(35) Destro, R.; Marsh, R. E.; Bianchi, R. A low-temperature (23 K) study of *L*-alanine. *J. Phys. Chem.* **1988**, *92*, 966−973.

(36) Capelli, S. C.; Bürgi, H.-B.; Mason, S. A.; Jayatilaka, D. Glycyl-*L*-alanine: a multi-temperature neutron study. *Acta Crystallogr., Sect. C: Struct. Chem.* **2014**, *70*, 949−952.

(37) Pichon-Pesme, V.; Lecomte, C.; Wiest, R.; Benard, M. Modeling fragments for the *ab initio* determination of electron density in polypeptides. An experimental and theoretical approach to the electron distribution in Leu-enkephalin trihydrate. *J. Am. Chem. Soc.* **1992**, *114*, 2713−2715.

(38) Sawaya, M. R.; Sambashivan, S.; Nelson, R.; Ivanova, M. I.; Sievers, S. A.; Apostol, M. I.; Thompson, M. J.; Balbirnie, M.; Wiltzius, J. J. W.; McFarlane, H. T.; Madsen, A. Ø.; Riekel, C.; Eisenberg, D. Atomic structures of amyloid cross-β spines reveal varied steric zippers. *Nature* **2007**, *447*, 453−457.

(39) Jelsch, C.; Teeter, M. M.; Lamzin, V.; Pichon-Pesme, V.; Blessing, R. H.; Lecomte, C. Accurate protein crystallography at ultra-high resolution: valence electron distribution in crambin. *Proc. Natl. Acad. Sci. U. S. A.* **2000**, *97*, 3171−3176.

(40) Allen, F. H.; Bruno, I. J. Bond lengths in organic and metal-organic compounds revisited: X-H bond lengths from neutron diffraction data. *Acta Crystallogr., Sect. B: Struct. Sci.* **2010**, *66*, 380−386.

(41) Turner, M. J.; Grabowsky, S.; Jayatilaka, D.; Spackman, M. Accurate and Efficient Model Energies for Exploring Intermolecular Interactions in Molecular Crystals. *J. Phys. Chem. Lett.* **2014**, *5*, 4249−4255.

(42) Bučinský, L.; Jayatilaka, D.; Grabowsky, S. Importance of Relativistic Effects and Electron Correlation in Structure Factors and Electron Density of Diphenyl Mercury and Triphenyl Bismuth. *J. Phys. Chem. A* **2016**, *120*, 6650−6669.

(43) Kohout, M. A measure of electron localizability. *Int. J. Quantum Chem.* **2004**, *97*, 651−658.

(44) Johnson, E. R.; Keinan, S.; Mori-Sánchez, P.; Contreras-García, J.; Cohen, A. J.; Yang, W. Revealing noncovalent interactions. *J. Am. Chem. Soc.* **2010**, *132*, 6498−6506.

(45) Arias-Olivares, D.; Wieduwilt, E. K.; Contreras-García, J.; Genoni, A. NCI-ELMO: A New Method To Quickly and Accurately Detect Noncovalent Interactions in Biosystems. *J. Chem. Theory Comput.* **2019**, DOI: 10.1021/acs.jctc.9b00658.

(46) Morra, G.; Genoni, A.; Neves, M. A. C.; Merz, K. M., Jr.; Colombo, G. Molecular Recognition and Drug-Lead identification: What Can Molecular Simulations Tell Us? *Curr. Med. Chem.* **2010**, *17*, 25−41.

(47) Genoni, A.; Pennati, M.; Morra, G.; Zaffaroni, N.; Colombo, G. Ligand selection from the analysis of protein conformational substates: new leads targeting the N-terminal domain of Hsp90. *RSC Adv.* **2012**, *2*, 4268−4282.

(48) Ferraro, M.; D'Annessa, I.; Moroni, E.; Morra, G.; Paladino, A.; Rinaldi, S.; Compostella, F.; Colombo, G. Allosteric Modulators of HSP90 and HSP70: Dynamics Meets Function through Structure-Based Drug Design. *J. Med. Chem.* **2019**, *62*, 60−87.

(49) Macetti, G.; Genoni, A. Quantum Mechanics/Extremely Localized Molecular Orbital Method: a Fully Quantum Mechanical Embedding Approach for Macromolecules. *J. Phys. Chem. A* **2019**, DOI: 10.1021/acs.jpca.9b08882.

(50) Jayatilaka, D.; Grimwood, D. J. Tonto: A Fortran Based Object-Oriented System for Quantum Chemistry and Crystallography. In *Computational Science − ICCS 2003*; Sloot, P. M. A., Abramson, D., Bogdanov, A. V., Dongarra, J. J., Zomaya, A. Y., Gorbachev, Y. E., Eds.; Springer-Verlag: Berlin & Heidelberg, 2003; Chapter 4, pp 142−151.

(51) Philipp, D. M.; Friesner, R. A. Mixed Ab Initio QM/MM Modeling Using Frozen Orbitals and Tests with Alanine Dipeptide and Tetrapeptide. *J. Comput. Chem.* **1999**, *20*, 1468−1494.

(52) Piltz, R. O.; Edwards, A. First results from the KOALA neutron Laue instrument. *Acta Crystallogr., Sect. A: Found. Crystallogr.* **2008**, *64*, C187.

(53) Spek, A. L. PLATON SQUEEZE: a tool for the calculation of the disordered solvent contribution to the calculated structure factors. *Acta Crystallogr., Sect. C: Struct. Chem.* **2015**, *71*, 9−18.

(54) Buergi, H. B.; Fischer, E.; Kunz, R. W.; Parvez, M.; Pregosin, P. S. Correlation between NMR coupling constants and molecular structure. Synthesis and ^{31}P NMR measurements of [HgX_2(*cis*-$Ph_2PCH = CHPPh_2$)] and x-ray crystal structures of [$HgBr_2$(cis-$Ph_2PCH = CHPPh_2$)], [$Hg(NO_3)_2(PPh_3)_2$], and [$Hg(CN)_2(PPh_3)_2$]. *Inorg. Chem.* **1982**, *21*, 1246−1256.

Chapter 5

Applying Quantum Crystallography for Rational Drug Design

The tools that the field of Quantum Crystallography provides showed to be capable of describing accurately the bonding, the environment and the behaviour of molecules and wavefunctions in the solid state. It was proven in previous publications and work of this thesis (see [92] and sections 4.2 & 4.1), that the procedure of XWR is capable of obtaining physical effects about, electron correlation, but also the polarization of a molecule in the environment from structure factors. As shown in section 4.1.4 this can also be achieved from the experimental diffraction data. Also, it was observed that if one uses a level of theory describing the effects in the *ab initio* wavefunction it can be expected that the wavefunction mimics the behavior of the molecule in the crystal better than a pure HF wavefunction. Therefore the HAR step in the following sections was always carried out using a DFT functional and a field of cluster charges to obtain the best possible geometry assuming we used the best possible wavefunction during the refinement. Then, the XCW step was carried out using a level of theory with the least amount of assumptions to retrieve all physical effects from the experimental data.

In the following investigations it will be studied whether these insights can be used to extrapolate the molecular properties observed in the crystalline phase into other environments, for example biological systems. To test whether the correlation of *Klebe* [274] can be extended from structural correlation towards electron density, polarization and bonding properties, the method of XWR will be used in comparison to QM/MM methods, to provide insight into the molecular wavefunction of ligands in biological and other environments.

5.1 *Publication:* Similarities and Differences between Crystal and Enzyme Environmental Effects on the Electron Density of Drug Molecules

The work on this publication started during my Master's thesis, which included the concept of bond scaled densities and quantification of differences through RRS values and integrated difference densities. The MD simulations and the QM/MM calculations were part of my Master's thesis. Erna K. Wieduwilt prepared the multipole model for the crystal structure data. The theoretical structure factors were calculated by me using *Denprop* and modelled in the same multipole model. During this PhD thesis, the HAR was performed using the new *Olex2-NoSpherA2* approach with subsequent XCW. The first draft of the publication was written by me. Simon Grabowsky rewrote the complete introduction and improved the discussion of results in several iterations. All Figures and Tables were generated by me. The synthesis and collection of X-ray dataset were carried out by Ming W. Shi and Simon Grabowsky, among others.

The results of this publication show a very promising trend that the crystal and protein electron density of a molecular wavefunction in crystalline and enzyme environments are the least different. The next best model, except for including an explicit neighbourhood, is the use of a continuum solvation model. This test case confirms and quantifies the claim of correlation between crystal and protein bound molecular properties (see section 1.4.2 and [274]). Since a solvation model seems to be a good approximation for the polarization in a crystal, it might have interesting implications for application in quantum crystallography for the future, if it was implemented in HAR.

However, this work also shows that radiation damage and problems that are due to measurement deficiencies can significantly damage the XCW procedure. This is directly shown by the interaction density and ESP of models CC (an *ab initio* wavefunction with cluster charges and dipoles) and XCW, where the density is basically trying to accommodate damages done to the crystal during the measurement. This highlights why HAR can even perform nice refinements with routine datasets. The theoretical model used to describe the diffraction pattern is *a priori*, while the XCW can become ill defined in the case of insufficient data quality. The experiment might not only provide useful information, but also physical effects of a diffraction experiment, that should not be introduced into the wavefunction.

To obtain the best results from the experimental data, a procedure is required which would compensate the effect of radiation on the intensities of measured data without contaminating the data with too many assumptions. So far, such a procedure is not reported to the best of my knowledge. Another possible treatment might include the introduction of decay-parameters in the model to compensate for the effect from the model perspective. This would work best if the diffraction data file contained either the dose that the crystal was already exposed to or the timing in the experiment, if a constant flux of X-rays is assumed. This model would require a deeper understanding to model the physical effect of the radiation damage, that might not always be available *a priori*. Also, changes in the absorption processes of the sample might be observed, if for example the oxidation state of elements in the crystal are changed due to the radiation, or the density changes, because some atoms might leave the crystal due to ionization. This would require much deeper understanding and, most likely, so far unfeasably complex models of the radiation damage.

Reprinted with permission from *Chemistry European Journal* **2020**. Copyright © 2020. DOI. 10.1002/chem.202003978.

Chemistry
A European
Journal

Chemistry
Europe
European Chemical
Societies Publishing

Accepted Article

Title: Similarities and differences between crystal and enzyme environmental effects on the electron density of drug molecules

Authors: Florian Kleemiss, Erna K. Wieduwilt, Emanuel Hupf, Ming W. Shi, Scott G. Stewart, Dylan Jayatilaka, Michael J. Turner, Kunihisa Sugimoto, Eiji Nishibori, Tanja Schirmeister, Thomas C. Schmidt, Bernd Engels, and Simon Grabowsky

To be cited as: *Chem. Eur. J.* 10.1002/chem.202003978

Link to VoR: https://doi.org/10.1002/chem.202003978

WILEY-VCH

WILEY-VCH

FULL PAPER

Similarities and differences between crystal and enzyme environmental effects on the electron density of drug molecules

Florian Kleemiss[a,b], Erna K. Wieduwilt[a,c], Emanuel Hupf[a], Ming W. Shi[d], Scott G. Stewart[d], Dylan Jayatilaka[d], Michael J. Turner[d], Kunihisa Sugimoto[e,f], Eiji Nishibori[g], Tanja Schirmeister[h], Thomas C. Schmidt[i], Bernd Engels[i], Simon Grabowsky*[a,b]

Abstract: The crystal interaction density is generally assumed to be a suitable measure of the polarization of a low-molecular weight ligand inside an enzyme, but this approximation has seldomly been tested and has never been quantified before. In this study, we compare the crystal interaction density and the interaction electrostatic potential for a model compound of loxistatin acid (E64c) with those inside cathepsin B, in solution and in vacuum. We apply QM/MM calculations and experimental quantum crystallography to show that the crystal interaction density is indeed very similar to the enzyme interaction density. Less than 0.1e are shifted between these two environments in total. However, this has non-negligible consequences for derived properties.

Introduction

Molecular recognition between an enzyme and a low-molecular weight ligand is a key feature in mode of action studies for biologically active molecules, and in consequence the most important factor in drug design.[1] The main components of molecular recognition are steric and electrostatic complementarity between the enzyme pocket and the active molecule. If the recognition process is simplified from the induced-fit theory to the classic key-lock mechanism,[2] small-molecule crystal structures can be used to approximate the correct three-dimensional shape of the active molecule in the biological environment. In both a biological as well as a crystalline environment, the small molecule will conformationally adapt its shape to the prevailing intermolecular binding forces, so that the resulting bound state reflects both its inherent flexibility and the environment.[3,4] Pascard states that “in numerous cases, there is a nearly perfect correlation between small-molecule structural results, and the observed binding in receptor–substrate complexes”.[5] Such similarities, investigated many times at the geometrical level,[6] have led to the development of the method of composite crystal-field environments by Klebe[7] and SuperStar[8] as part of the Cambridge Structural Database suite of software.

Electrostatic complementarity between enzyme binding site and active molecule is an aspect that goes beyond geometry and molecular conformation since the electrostatic potential is inherently related to the electron density distribution of each partner.[9] Electron densities of biologically active small molecules can computationally be estimated in different environments (isolated state, solution, crystalline state) or modeled from experimental X-ray diffraction structure factors.[10] Experimental electron-density determinations of crystals of low-molecular weight enzyme ligands give a detailed insight into intermolecular interactions which are relevant for the biological recognition process such as electrostatic forces, hydrogen bonding or van-der-Waals interactions.[11] Consequently, the numerous experimental electron-density investigations of biologically active compounds were justified by assuming that the polarization of the molecule under scrutiny reflects the polarization in the enzyme.[12] However, this underlying assumption has only rarely been investigated,[10,13] and, to the best of our knowledge, never been quantified. In this study, we quantify the extent of similarity between the electron-density distributions of a model compound of the drug E64c as computationally determined in vacuum, in aqueous solution, in the corresponding cathepsin-B enzyme complex and in the crystal structure of the pure molecule. The electron-density distribution of the compound in its crystal structure was also determined from a low-temperature high-resolution synchrotron X-ray diffraction study.

The epoxysuccinyl peptide loxistatin acid (E64c) is a derivative of the natural product E64[14] and a potent inhibitor of

---

[a] F. Kleemiss, E. K. Wieduwilt, Dr. E. Hupf, PD Dr. S. Grabowsky
Institute of inorganic Chemistry and Crystallography
University of Bremen
Bremen, Germany
[b] F. Kleemiss, PD Dr. S., Grabowsky
Department of Chemistry and Biochemistry
University of Bern
Bern, Switzerland
E-Mail: simon.grabowsky@dcb.unibe.ch
[c] E. K. Wieduwilt
Laboratory of Theoretical Physics and Chemistry
University of Lorraine
Metz, France
[d] M. W. Shi (deceased), A/Prof. S. C. Stewart, Prof. D. Jayatilaka, Dr. M. J. Turner
School of Molecular Sciences
University of Western Australia
Perth, Australia
[e] A/Prof. K. Sugimoto
Japan Synchrotron Radiation Research Institute
SPring-8
Hyogo, Japan
[f] A/Prof. K. Sugimoto
Institute for Integrated Cell-Materials Sciences
Kyoto University
Kyoto, Japan
[g] Prof. E. Nishibori
Tsukuba Research Center for Energy Materials Science
University of Tsukuba,
Tsukuba, Japan
[h] Prof. Dr. T. Schirmeister
Institute of Pharmaceutical and Biomedical Sciences
Johannes-Gutenberg University Mainz
Mainz, Germany
[i] Dr. T. C. Schmidt, Prof. Dr. B. Engels
Institute for Physical and Theoretical Chemistry
Julius-Maximilians-University Würzburg
Würzburg, Germany

Supporting information and the ORCID identification number(s) for the authors of this article can be found under:
https://doi.org/10.1002/chem.xxxx

papain-like cysteine proteases (CAC1 enzymes).[15] E64c irreversibly inhibits proteases through nucleophilic epoxide ring opening and formation of a new C-S covalent bond with the respective cysteine thiol group. This mode of action has been supported by mechanistic studies[16] based on the E64c–cathepsin B complex crystal structure[17]. Furthermore, the crystal structure of the pure E64c ligand alone was determined in 2015.[4] It was envisaged that this ligand would be ideal for the anticipated experimental study examining the similarities and differences between the polarization of the electron density of a drug molecule in its crystal structure to that in an enzyme since so much computational information about E64c is available.[13,14] However, the four independent molecules of E64c in its crystal structure are related by *pseudo*-symmetry and are heavily disordered, which makes them unsuitable for experimental electron-density determinations.[4] In turn, the E64c model compounds that were in fact suitable for experimental ED investigations[18] do not fit into any of the known cysteine proteases targeted by E64c (*in vivo* or *in silico*) preventing a comparison between the crystal and enzyme environments. To overcome these challenges, we recently devised an alternative strategy which first involved synthesis of the model molecule (2S,3S)-3-aminocarbonyl-oxirane-2-carboxylic acid (**1**H).[19] As in E64c, this new substrate contains both the required electrophilic epoxide moiety for the reaction with cysteine and the carboxyl anchor group for initial binding into the enzyme pocket.

At physiological pH-values, the carboxylic acid group in E64c is deprotonated and hence binds to cathepsin B as a carboxylate anion. Therefore, in the present study we also deprotonated **1**H yielding the anion **1** (Figure 1a), which crystallizes as a potassium salt with one molecule of co-crystallized water (**1**K·H_2O, Figure 1b). Water is abundant in the E64c–cathepsin B complex, too, and the K^+ ion in the crystal structure replaces the histidinium residue of cathepsin B in the complex, so that the type of intermolecular interactions in the crystal studied here can be expected to mirror those inside cathepsin B (section 1 of the Results and Discussion). A single crystal of **1**K·H_2O was measured to high resolution using synchrotron radiation at BL02B1, SPring-8, and the electron-density distribution was subsequently modelled experimentally. For this purpose, we employed X-ray Wavefunction Refinement (XWR,[20]), a novel technique combining Hirshfeld Atom Refinement (HAR,[21]) and X-ray constrained wavefunction (XCW) fitting.[22,23] *In silico*, the small anion **1** fits into the binding pocket of cathepsin B, so that its electron density could be compared between the following environments: in the crystal structure (from XWR, model ***X***; from QM/MM calculations, model ***C*** for *crystal*; embedded in a self-consistent field of cluster point charges, model ***CC*** for *cluster charges*), inside the enzyme cathepsin B (from QM/MM calculations, model ***P*** for *protein*), in solution (COSMO solvation model ***S***) and in vacuum (model ***G*** for *gas phase*).

Intermolecular interactions between molecules lead to a polarization of each molecular electron density. In crystallography, the *interaction density* is defined as the difference between the electron density of a molecule in a crystal and the electron density of the same molecule with identical geometry in an assembly of non-interacting molecules (*in-vacuo* model). In biology, the difference between the molecule bonded to the active site and the molecule solved in water outside the active site is decisive. Nevertheless, the interaction density of biologically active small molecular compounds has often been used as a first approximation to the polarization of the ligand in the enzyme. E.g., Dittrich and Matta argue that the “interaction density can be seen as an idealized situation that is analogous to drug–receptor interactions and the redistribution of electron density of a drug molecule in the active site.”[24]

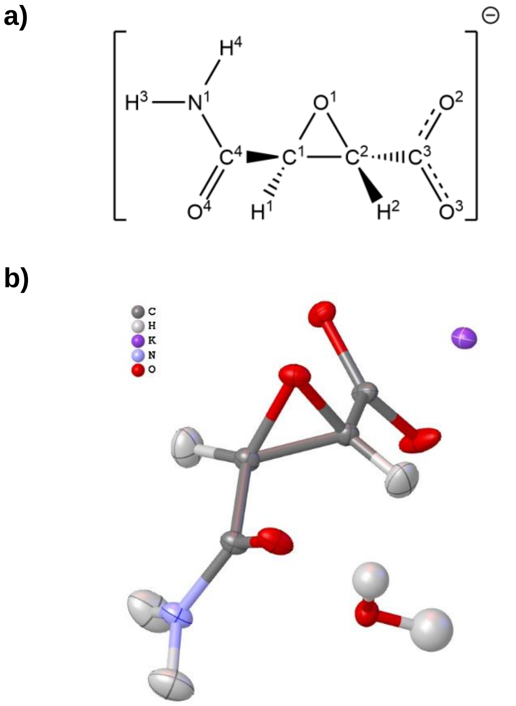


Figure 1. a) Illustration and atomic numbering of the *(2S,3S)*-3-aminocarbonyloxirane-2-carboxylate anion (**1**). b) Refined structure of the formula unit **1**K·H_2O after HAR. All hydrogen atom positions and the isotropic displacement parameters for the water hydrogen atoms were freely refined. Anisotropic hydrogen displacement parameters in **1** were estimated with Shade2[31] and fixed during the HAR. All displacement parameters are drawn at a 50% probability level with the software Olex2[32]. The refined intramolecular geometry is as follows: C1-O1=1.4221(6), C2-O1=1.4306(6), C1-C2=1.4761(7), C2-C3=1.5204(7), C3-O2=1.2696(6), C3-O3=1.2351(7), C1-C4=1.5116(6), C4-O4=1.2286(6), C4-N1=1.3372(6), C1-H1=1.064(17), C2-H2=1.070(18), N1-H3=1.087(19), N1-H4=1.030(19), O^{H_2O}-H=0.95(2)/0.96(2) Å; C1-O1-C2=62.32(3), C1-C2-O1=58.56(3), C2-C1-O1=59.12(3), O2-C3-O3=126.13(5), H-O^{H_2O}-H=108.0(18) °.

It is arguable if the magnitude of the interaction density is inside or outside experimental errors for charge-density experiments, which means it is unclear if the interaction density can be measured.[25] Most of the discussion in the past was centered around the accuracy of the multipole model for extracting interaction-density characteristics.[26] Alternatively, the interaction density was calculated based on periodic-boundary theoretical computations.[27] Only very recently, an alternative approach to the derivation of the interaction density was discussed, namely the comparison of the non-fitted and maximally fitted molecular wavefunction according to the X-ray constrained wavefunction fitting technique.[28] The non-fitted wavefunction produces *the in-vacuo* electron density of the molecule under examination, whereas the fitted wavefunction includes the polarization of the molecule by the crystalline environment via the measured X-ray structure factors. This new method was described mainly for an application to theoretically calculated X-ray structure factors. In this work, we test this method against experimental structure factors for the compound **1**K·H_2O (section 2 of the Results and Discussion).

Although the interaction density directly reflects the polarization of the molecule by its environment, for the molecular recognition process the electrostatic potential is the crucial property. Electron density and electrostatic potential are fundamental properties of a wavefunction and in this study both are calculated directly from the (experimentally constrained) wavefunctions. Beyond this relationship, there are also interesting similarities and differences between the topologies of their scalar fields and how these are interpreted with respect to chemistry and biology.[29] However, to the best of our knowledge, the electrostatic-potential analogue to the interaction density has never been explored before. We refer to this difference between *in-vacuo* and crystal electrostatic potential as *interaction electrostatic potential* (or simply *interaction ESP*) since the term interaction potential has a different meaning in the framework of QM/MM calculations.[30]

As discussed above, *interaction density* (and consequently also *interaction ESP*) are crystallographic terms, denoting the difference between the crystalline and the *in-vacuo* states at identical geometry. However, for the simplicity of the argument, here we generalize both terms to denote the difference between the polarized (crystal **X** or **C** or **CC**, enzyme **P**, solution **S**) and unpolarized (*in-vacuo*, gas phase, **G**) states at identical geometry (section 2 of the Results and Discussion). In addition, we have developed a technique to compare the electron densities of different optimized molecular geometries, which will be introduced in section 3 of the Results and Discussion. In section 4, we compare derived properties such as the Laplacian of the electron density and atomic charges in different environments.

Results and Discussion

1. Comparison of intermolecular interactions

Initially, a comparison between the intermolecular interaction network of **1** inside cathepsin B and that inside the crystal structure of **1**K·H_2O (Figure 2) is discussed. For that purpose, molecular dynamics (MD) input was prepared by taking the E64c–cathepsin B complex crystal structure[17] and pruning off all atoms of E64c that are not part of the simplified **1**. After initial equilibration in a water box, molecular dynamics were simulated over 5 ns. The interaction of the carboxylate group with His-199 and Cys-29 (Figure 2, left) is the electrostatic anchor that keeps **1** closely bonded inside the enzyme pocket during the entire time of the simulation (Figure 3, O_3-$HN_{His+-199}$). The contact with carboxylate atom O2 importantly keeps Cys-29 close to the epoxide ring, eventually resulting in an interaction between the sulfur atom of Cys-29 and the epoxide carbon atom 1, which prepares the ring-opening reaction of E64c to irreversibly form a covalent C-S bond and inactivate the cysteine protease.[16] Figure 3 shows that the C_1-S_{Cys-29} contact remains stable throughout the MD simulation with **1** in cathepsin B implying that the model compound used here indeed represents the major characteristics of E64c as it would act in the same bonding environment. However, the amine group of **1** points into the artificially produced cavity, produced by shortening of the ligand compared to E64c, which is now filled with water molecules. Figure 2, left, shows that the nearest enzyme residues are Gly-27 and Gly-74, but as can be seen in Figure 3, these contacts are rather long, giving rise to conformationally flexible and rotating amide groups.

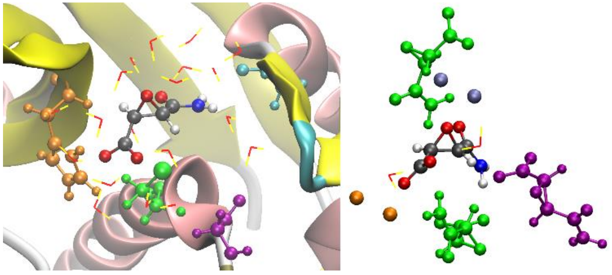


Figure 2. Visualization of close contacts of **1** in **P** (left) and **X** (right). Coloring according to interaction type. Orange: His^+-199/ K^+ around carboxylate; green: Cys-29/ **1** (symmetry-generated) around epoxide and carboxylate; purple: Gly-27/ **1** (symmetry-generated) around amine; cyan/blue: Gly-74/ K^+ around carbonyl in amide group. Water is depicted as red-and-yellow sticks.

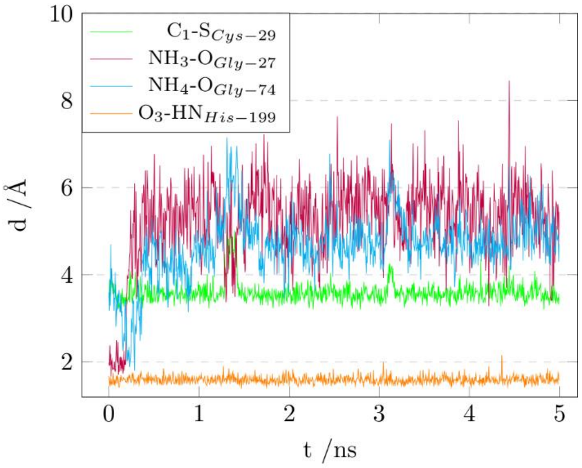


Figure 3. Plot of the distance between **1** and the four closest amino-acid residues of the cathepsin-B pocket (as visualized in Figure 2, left) during the 5 ns MD simulation of **1** inside cathepsin B.

Geometry snapshots were taken every 250 ps along the MD simulation between 4 and 5 ns. Starting from these, a total of 4 different QM/MM geometry optimizations were carried out. The resulting geometries are similar but not identical due to the multi-minima nature of protein conformations. Table 1, left column, shows that the highest deviations for the various intermolecular contacts of the 4 final geometries vary between 0.16 and 0.01 Å, so that it is meaningful to report and interpret their average for a direct comparison between the protein and crystal environments (yielding model **P** for the following sections).

For every hydrogen bonding donor or acceptor, there is either a hydrogen bond or an electrostatic contact in both these environments. The carboxylate group (atoms O2 and O3) forms one hydrogen bond with water as acceptor and one contact dominated by electrostatics (either positive histidinium ring or positive K^+ ion) each. The amine group forms two hydrogen bonds to water inside the enzyme pocket, and two hydrogen bonds to a symmetry-generated carboxylate atom O2 in **1**K·H_2O. The epoxide oxygen atom O1 forms a hydrogen bond or an electrostatic contact to K^+, respectively, with the same distance of

Table 1. Closest contacts with the environment inside cathepsin B (left) and inside the crystal structure of **1**K·H_2O (right). For hydrogen bonds D-H…A, the distances refer to the donor-acceptor distances D…A. For the atom labelling, see Figure 1 a).

1 in cathepsin B		**1**K·H_2O	
Contact (**P**)	Distance /Å[a]	Contact (**X**)	Distance /Å[b]
$O^1\cdots H\text{-}O^{H_2O}$	2.89 (4)	$O^1\cdots K^+$	2.9092 (4)
$N^1\text{-}H^3\cdots O^{H_2O}$	2.76 (7)	$N^1\text{-}H^3\cdots O^2$	3.0157 (6)
$N^1\text{-}H^4\cdots O^{H_2O}$	2.98 (16)	$N^1\text{-}H^4\cdots O^2$	3.0443 (6)
$O^2\cdots H\text{-}O^{H_2O}$	2.69 (4)	$O^2\cdots H\text{-}O^{H_2O}$	2.7468 (6)
$O^3\cdots H\text{-}N^{His+\text{-}199}$	2.69 (3)	$O^3\cdots K^+$	2.6210 (5)
$O^4\cdots H\text{-}O^{H_2O}$	2.65 (1)	$O^4\cdots K^+$	2.6925 (4)

[a] The distances are the average of 4 QM/MM optimized geometries (see Experimental and Computational Section). The numbers in brackets are the sample standard deviations in the last shown digit(s). [b] The distances are the crystallographically refined geometries from the XWR procedure. The numbers in brackets are the standard uncertainties in the last shown digit from the refinement.

2.9 Å. Remarkably, the longest of the contacts coincide in both environments, and the shortest one in **1**K·H_2O coincides with the second-shortest one in the enzyme, which is, however, the most important one, namely the electrostatic anchor identified in Figure 3.

In summary, a qualitative comparison of the intermolecular interaction networks in both crystal and enzyme environments shows some systematic similarities that lead us to believe that a comparison of the polarization of the electron density in both environments will be meaningful. However, Figure 2 also shows that the conformations of the carboxylate and the amide groups in both environments are significantly different, so that a direct comparison of the electron densities belonging to individual bonds will be difficult. This problem will be tackled in section 3, whereas section 2 deals with the interaction densities and interaction electrostatic potentials of the entire anion **1**.

2. Interaction density and interaction electrostatic potential

The interaction densities (Figure 4) and interaction electrostatic potentials (ESPs, Figure 5) were calculated according to the ideas discussed in the Introduction. They represent the differences between anion **1** in its polarized (crystal **X** or **C** or **CC**, enzyme/protein **P**, solution **S**) and unpolarized (*in-vacuo*, gas phase, **G**) states at identical geometry. In detail, Figures 4 and 5 depict the following model differences:

4 a)/ 5 b) and c): model **P** *minus* model **G** at the QM/MM-optimized geometry of **P**;

4 b)/ 5 e) and f): model **C** *minus* model **G** at the QM/MM-optimized geometry of **C**;

4 c)/ 5 h) and i): polarization of **1** with a cluster of Hirshfeld-atom point charges and dipoles (model **CC**, cluster charges) *minus* model **G** at the experimental geometry from X-ray diffraction (**X**);

4 d)/ 5 k) and l): effect of X-ray constrained wavefunction fitting:[22,28] X-ray constrained model **X** *minus* model **G** at the experimental geometry from X-ray diffraction (**X**).

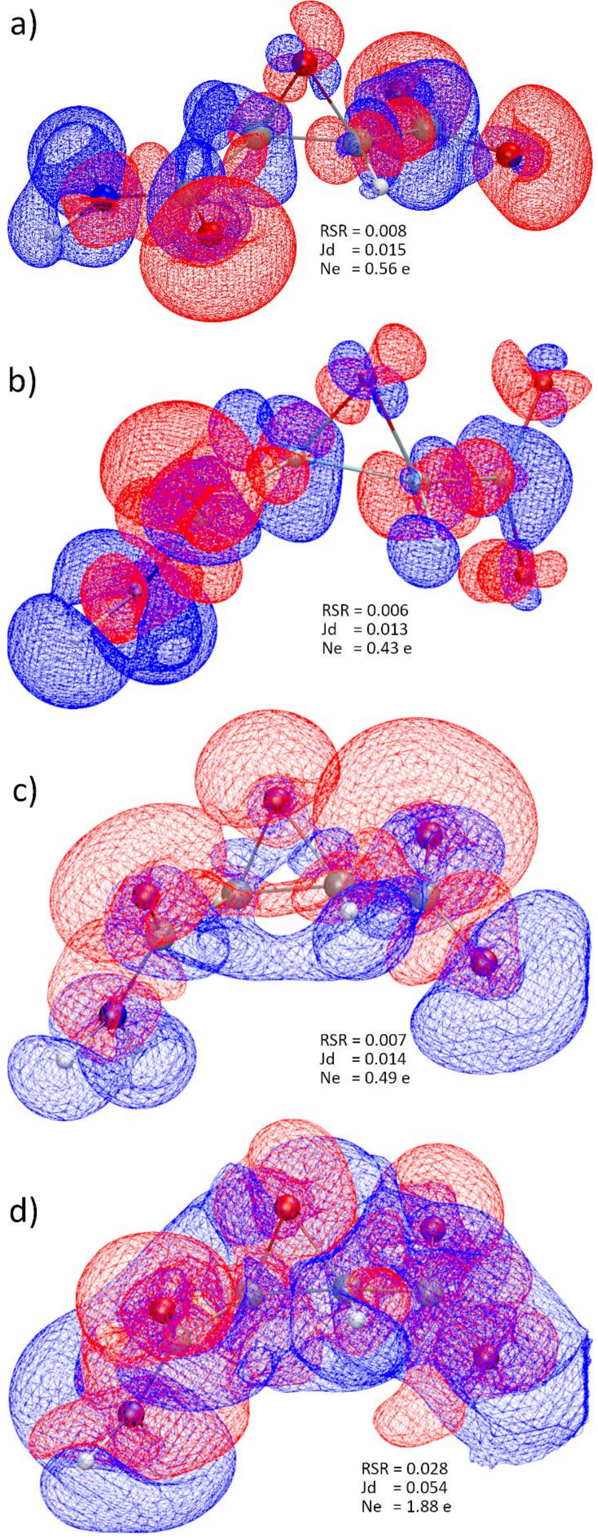


Figure 4. Isosurfaces of interaction densities at ±0.0067 e Å^{-3}. Red regions correspond to higher ED compared to the *in-vacuo* state, blue to lower ED. a) Protein env.: **P** minus **G** at geom. **P**; b) crystal env.: **C** minus **G** at geom. **C**; c) cluster charges: **CC** minus **G** at geom. **X**; d) X-ray constrained wavefunction: **X** minus **G** at geom. **X**. RSR = real-space R-value. Jd = Jaccord distance. Ne = integrated number of electrons in the difference grid file. All isosurface representations are generated with VMD.[36]

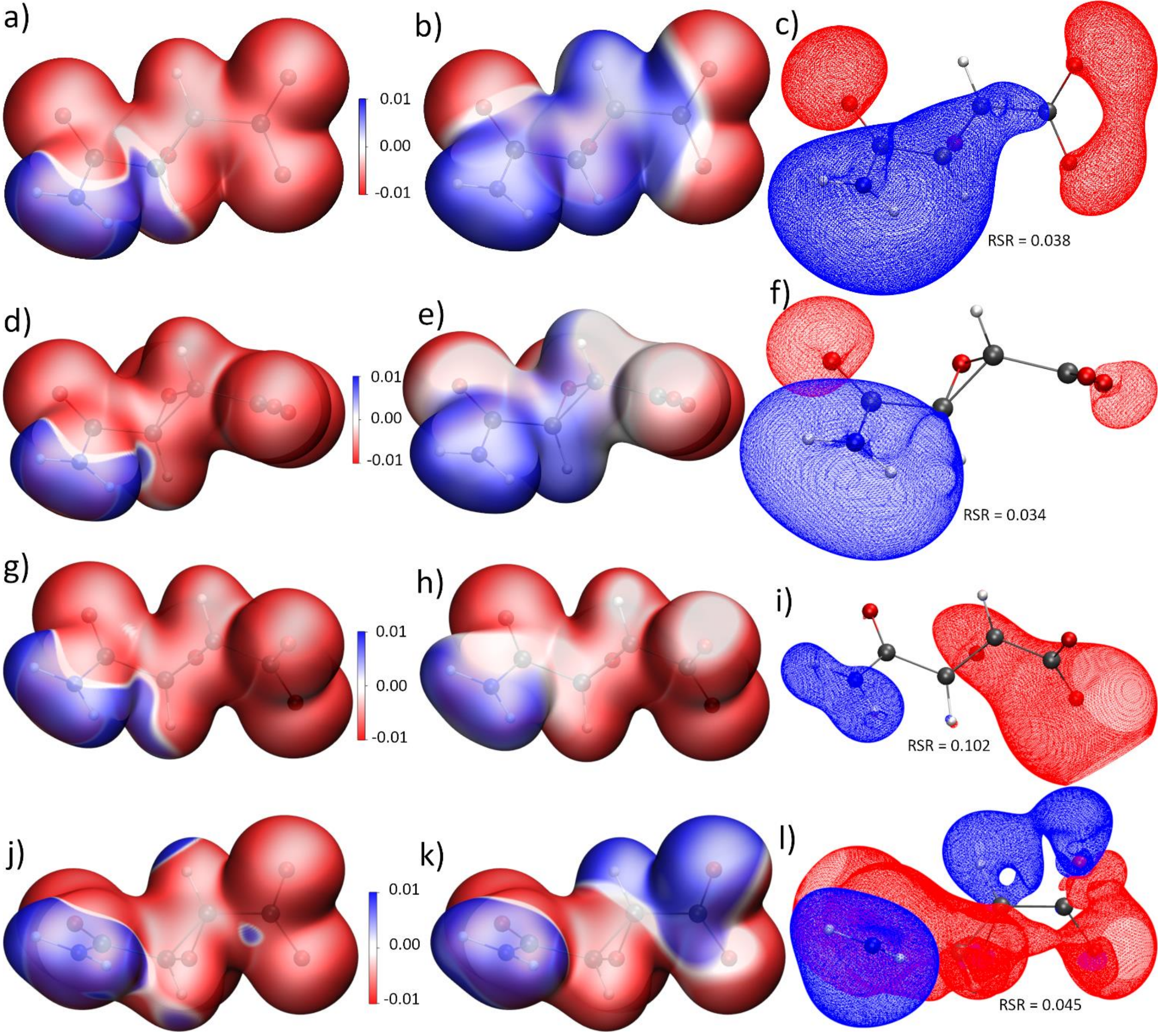


Figure 5. Left row (a,d,g,j): 0.01 a.u. isosurface of the ED with total ESP mapped onto it, capped at ± 0.01 *e* Å^{-1}. Middle row(b,e,h,k): 0.01 a.u. isosurface of the ED with interaction ESP mapped onto it, capped at ± 0.01 *e* Å^{-1}. Right row(c,f,i,l): Isosurfaces of the interaction ESP at ±0.028 *e* Å^{-1}, except for i) with +0.014 and -0.070 *e* Å^{-1}. Red corresponds to more negative regions compared to the *in-vacuo* state, blue to more positive regions (compare color scale). a) Protein env.: **P** at geom. **P**; b,c) Protein env.: **P** minus **G** at geom. **P**; d) crystal env.: **C** at geom. **C**; e,f) crystal env.: **C** minus **G** at geom. **C**; g) cluster charges: **CC** at geom. **X**; h,i) cluster charges: **CC** minus **G** at geom. **X**; j) X-ray constrained wavefunction: **X** at geom. **X**; k,l) X-ray constrained wavefunction: **X** minus **G** at geom. **X.** RSR = real-space R-value of the interaction ESP. (Units: 1 *e* Å^{-1} = 14.40 *V*.)

In the QM/MM calculations, only **1** was treated quantum-mechanically, but everything else including water and potassium ions was treated with molecular mechanics, so that the polarization features seen in Figures 4/5 a) and b) are purely caused by electrostatics, with limited dispersion and no covalent component. The calculations of Figures 4/5 c) and d), however, required the definition of at least the asymmetric crystal unit, which explicitly includes water and potassium ions. In the Experimental and Computational Section, we describe the methodology developed here to isolate only the electrostatic component of the interactions.

The interaction densities in Figures 4 a) and b) look visually similar. The similarity is even more obvious in the interaction ESP plots, comparing Figure 5 b) with e) and c) with f). Note that the conformations of the amide and carboxylate groups are different in crystal and enzyme environments, so that the lobes around the carboxylate group in Figure 5 f) need to be turned by 90° to match the shape of the lobes in Figure 5 c). This means that the same functional groups are qualitatively polarized in the same way in both environments **P** and **C**. For example, all oxygen lone pairs show red regions in the interaction densities and ESPs, which means that they accumulate electron density and become more negatively polarized in the **P** and **C** environments than in the *in-vacuo* state **G**. All hydrogen atoms are covered in blue surfaces, which means that they lose electron density relative to the isolated state. Both effects are due to interaction with neighboring noncovalent bonding partners, e.g. via hydrogen bonding, which is characterized by charge transfer.

In absolute numbers, the polarization effect on the ED and ESP is small; the difference values are low. For the interaction density, the maximum value in the grid file is ±0.1 $eÅ^{-3}$ compared to maximum values of up to 265 $eÅ^{-3}$ in the grid file of the total ED. For the interaction ESP, the maximum values in the grid file are ±0.08 $eÅ^{-1}$ compared to maximum values in a range of [-0.31,782] $eÅ^{-1}$ for the total ESP. A meaningful isovalue for the representation of the interaction densities is 0.007 e $Å^{-3}$, whereas it is around 0.03 e $Å^{-1}$ for the interaction ESP. The real-space R-values (RSR) and Jaccord distances (Jd) are still close to zero. (For a definition of these quantities, please see the Experimental and Computational section.) An integration of the absolute values of the interaction density inside the grid files (N_e = number of electrons) is a measure of the charge transfer caused by the polarization (Figures 4 a) and b)). A total of 0.56 electrons are shifted in the protein environment and 0.43 e in the crystal environment, indicating that, overall, **1** is slightly more polarized by the enzyme than the crystal. Definitions of RSR, Jd and N_e are given in the Experimental and Computational section.

Dittrich et al. suggested to use different point-charge and -dipole models to simulate the crystal field effect and produce the interaction density.[27c] Following this idea, we used a self-consistent Hirshfeld cluster-charge model as it is normally employed in Hirshfeld Atom Refinements.[21b,33] Figure 4 c) shows the effect of this field on the ED: qualitatively, the regions of accumulation and depletion are the same around oxygen and hydrogen atoms in comparison to the QM/MM results, and quantitatively, RSR, Jd and Ne values are in-between the respective values for the QM/MM results. However, the regions of ED polarization are wider and more diffuse. This has strong consequences on the ESP, see Figures 5 h) and i). The polarization of all the oxygen atoms by the surrounding positive point charges is much stronger than in the QM/MM models: the negative isovalue in Figure 5 i) is 2.5 times as high and the RSR value is three times as high. Hence, it is clear that the Hirshfeld cluster-charge model overestimates the polarization effect by the crystal field.

Finally, following an idea by Ernst et al. [28], we have employed XCW fitting against the experimental synchrotron X-ray structure factors of **1**$K{\cdot}H_2O$ to map the crystal field effect on **1**. The experiment contains more information, for the better or worse, than the quantum-mechanical ansatz used for the XCW. Beyond polarization via the crystal field, additional effects are electron correlation,[34] and, unfortunately, systematic errors, with radiation damage being the most important one here. These effects are also included into the X-ray constrained wavefunction, so that the result is not a pure interaction density or interaction ESP.[23] However, since a hybrid-DFT functional was used in the fitting, we assume that the major electron correlation effect is included already in the unperturbed wavefunction before fitting, so that the major effect presented here is indeed due to the crystal field and the experimental error.

Figure 4 d) shows that the effect of the XCW fitting on the ED is much more pronounced than that of the theoretical models where the polarization effect could be treated in an isolated way. The spread of the isosurfaces as well as the RSR and N_e values are much larger, showing that 1.88 e are shifted between the unperturbed and the fitted wavefunction, which is nearly four times more than in the QM/MM and point-charge models. However, qualitatively the experimental information is still meaningful as the polarization pattern in the interaction density is still the same (red at oxygen, blue at hydrogen, see also representations at different isovalues in the Supporting Information, Figures S8). In the interaction ESP (Figures 5 k,l)), the polarization of the carboxylate group is unsymmetrical, i.e. one oxygen atom becomes more and one less negatively polarized in the crystal environment, which is the only qualitative difference to the theoretical results in Figures 5 a) to i). The RSR value of 0.045 is close to those of the QM/MM calculations.

The high-resolution low-temperature synchrotron X-ray diffraction data of **1**$K{\cdot}H_2O$ used in this study were prone to radiation damage, which was corrected with a scaling procedure (see Supporting Information). It is conceptually difficult within XCW fitting to find out to which extent the data quality is affected, which will in turn affect the results presented here. Therefore, we decided to perform a multipole modeling according to Hansen and Coppens[35] on the experimental and on additional theoretically calculated structure factors with the sole purpose of separating model from data effects, which is discussed in detail in the Supporting Information. We find that significant discrepancies are caused by the model used to treat the experimental data and not by data inaccuracy alone.

In summary, qualitatively and quantitatively the polarization of the electron density and the ESP of **1** in crystal and enzyme environments is similar, with a slightly larger effect in the enzyme. Hirshfeld cluster charges overestimate the crystal field effect, whereas XCW fitting against the experimental structure factors captures the effect with a strong bias in the ED and a smaller bias in the ESP. Since every model for experimental X-ray data treatment, such as the XCW or the multipole model, includes theoretical assumptions, and since the data always consist of a convolution of many physical effects, we believe that the refinement of experimental data might be useful, but data can never simply be trusted, as also demonstrated in ref.[37] In this context, multipole-model derived electron densities were compared to the crystal electron densities calculated from QM/MM approaches at various different levels of theory in previous studies.[38-40] However, it remains unclear, here and in general, whether the level of data accuracy and the convolution of different physical effects in the data yields physically useful and meaningful derivations of the interaction density and ESP from the experiment.[25,26] Therefore, in the next section we concentrate on details of the difference densities from theory alone.

3. Bond-wise comparison of difference densities

One disadvantage of the interaction density is the reference to isolated *in-vacuo* systems. Therefore, we now use difference

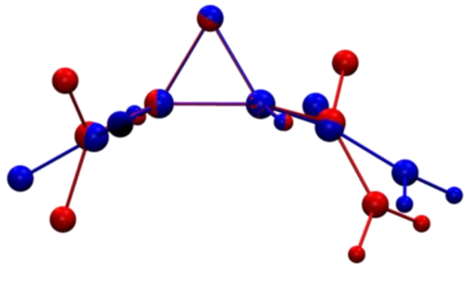


Figure 6. Superposition of the optimized molecular geometry of **1** in cathepsin B (model **P**, blue) and in the crystal of **1**$K{\cdot}H_2O$ (model **X**, red).

densities that use the protein environment as reference. However, a conceptual difficulty to overcome is that such comparisons are necessarily based on different molecular geometries. The molecular conformations of **1** obtained from geometry optimization in the crystal structure (QM/MM, model **C**), inside the enzyme cathepsin B (QM/MM, model **P**), in solution (COSMO, model **S**) and in vacuum (model **G**) are significantly different, as visualized in Figure 6 for the **C** and **P** geometries.

To account for changes of bond distances and angles, only bond-centered grids were analyzed that were chosen to be as small as possible and scaled to the respective bond lengths, i.e. the sample point separation was adjusted relative to the bond lengths (Figure 7, more details in the Experimental and Computational section). Following this approach, only minor contaminations of the difference densities in the corners of the boxes used to define the grids are observed.

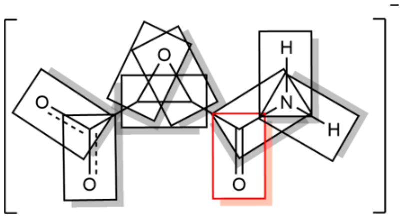


Figure 7. Visualization of bond-centered scaled grids and the selection of all bonds of **1** used for the analysis in section 3. The C4-O4 bond is highlighted, which is the first entry in Figures 8 and 9.

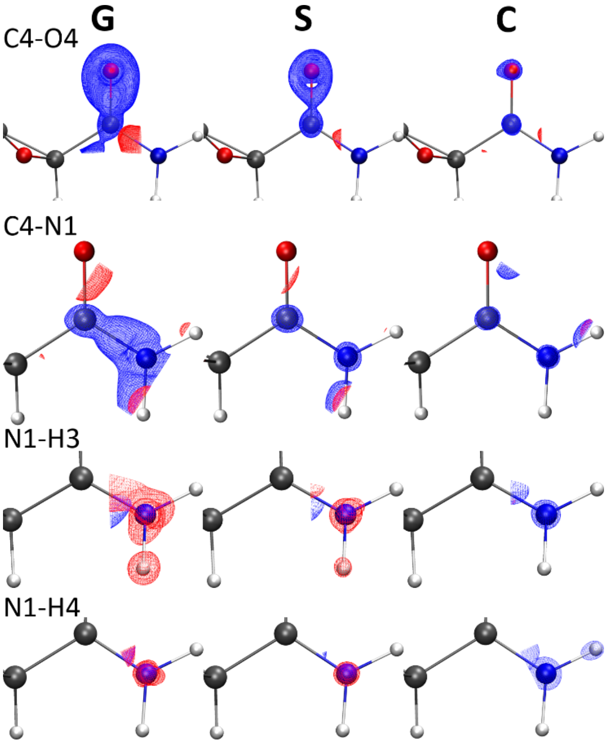


Figure 8. Isosurfaces of difference electron densities in bond-centered scaled grids at ±0.054 $e\,Å^{-3}$ (blue = positive, red = negative). Differences are shown for those four bonds where the effect is most pronounced (cf. Figure 9), and for the differences of the vacuum (**G**), solvent (**S**) and crystal (**C**) models always with respect to the enzyme (**P**).

It is expected that the effect of polarization by the environment for already inherently polarized bonds that interact with the environment is most pronounced. Whereas the carboxylate bonds C3-O2 and C3-O3 that carry the negative charge are less affected, Figure 8 shows the density differences for the most affected bonds C4-O4 and C4-N1. The discrepancy between models **G** and **P** is the largest, significantly reduced in the solvation model **S**, and nearly vanished in the crystal environment **C**. This implies that the crystal environment is indeed a good model of the larger enzyme environment. For the N-H bonds in Figure 8, the **G**-vs-**P** difference is always the most pronounced as for the C-N and C-O bonds. However, the trend between **S**-vs-**P** and **C**-vs-**P** is more ambiguous. In addition, for the N-H bonds the impurities in the corners of the boxes are higher.

Nevertheless, RSR values were calculated for every bond-centered difference grid file and graphically summed up in Figure 9 for the same differences as in Figure 8 (**G**/**S**/**C** vs. **P**), but in addition the K$^+$ counter-cation was explicitly included in the **G** and **S** calculations. K$^+$ is located in proximity to the carboxylate group replacing the His-199 group in the protein environment (Figure S9 in the Supporting Information). This first polarization by K$^+$ already reduces the differences between the **G** and the **P** environments, as it introduces a first polarizing field acting upon the molecule. This shows that the COSMO solvation model **S** produces already a good approximation to the polarization present in the enzyme[13a], so the influence of K$^+$ becomes negligible. However, the crystal environment is the best approximation to the protein environment, and if the bond N1-H4 was not an outlier, the effect would be visually even more compelling. In total, the similarity improves from an RSR sum of 0.171 for model **G** (w/o K) to 0.072 for model **C**, which is a drop of 0.1 RSR points or 58 %.

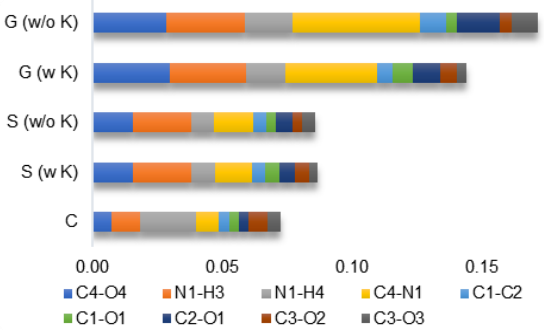


Figure 9. Sum of the RSR values for all bond-centered difference density grid files as visualized in Figure 8 for the four most affected bonds. The differences of the vacuum model (**G**) with or without K$^+$ counter-cation, the solvent model (**S**) with or without K$^+$ counter-cation and the crystal model (**C**) are always calculated with respect to the enzyme model (**P**). See also Figure S10 in the SI.

4. Comparison of derived properties

As the second derivative of the electron density, the Laplacian is very susceptible to any method or model change. Therefore, we decided to plot it along the C4-N1 bond, which is the one most affected by the environmental influences as shown in Figures 8 and 9. Figure 10 shows the progression of the Laplacian along this bond for all different models including the Hirshfeld cluster-

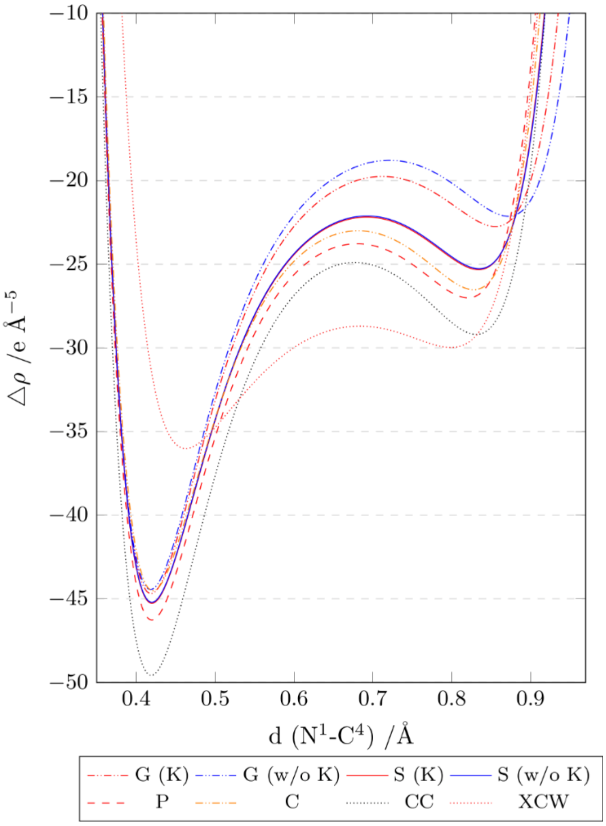


Figure 10. Laplacian of ED along the N1-C4 bond (plotted with the N atom at position 0 Å, left) for the vacuum model (**G**) with or without K^+ counter-cation, the solvent model (**S**) with or without K^+ counter-cation, the crystal model (**C**), the enzyme model (**P**), the Hirshfeld cluster-charge model (**CC**) and the XCW fitting model (**X**). Values were calculated using Multiwfn.[41]

charge model (**CC**) and the XCW-fitted model (**X**). The minima at ca. 0.42 and 0.85 Å are the valence-shell charge concentrations (VSCCs) inside this polar covalent bond. With the **P**-results serving as the reference as in the previous section, the deeper VSCC at the nitrogen atom is well described by all theoretical models except the **CC** model, which overestimates the charge concentration. The experimental data in the XCW fitting model significantly deviate from all other results, producing a shift of the nitrogen VSCC by about 10 eÅ^{-5} and 0.2 Å. This effect was previously observed and reported for the multipole model.[38,39]

The deviations of all models relative to the model **P** are best visible in the middle of the bond towards the carbon atom VSCC. Model **G** differs the most from the protein model **P**, which is slightly improved when the counter-cation K^+ is accounted for. The results are significantly closer to the reference model **P** for both solvation models (**S**), but the closest match exists between the crystal (**C**) and protein model. These findings agree with those in Figure 9 for the difference densities. The model **CC** has the most negative Laplacian value (highest electron concentration) of the theoretical models everywhere along the bond, but it is still closer to the models **P** and **C** than the model **G**. In contrast to the model **CC**, whereas the XCW fitting model shows the least negative Laplacian value for the nitrogen VSCC, it shows the most negative Laplacian value in the middle of the bond, deviating significantly from all other models. As discussed in section 2, it is difficult to know whether these deviations are caused by data inaccuracy or the convolution of many different physical effects – not only the crystal field effect whose treatment is the only difference between all the theoretical models. A comparison with Figure S7 in the Supporting Information shows, however, that the multipole model using both the experimental and theoretically generated structure factors leads to hugely different results of the Laplacian in this polar covalent C-N bond, and relative to that result the XCW fitting model agrees well with the theoretical models (compare a related discussion in ref.[42]).

The atomic charges (Table 2, here calculated according to Bader's QTAIM[43]) are directly related to the electrostatic potentials. All the oxygen atoms are most negatively charged in the protein environment, losing negative charge from models **C** toward model **S**, and being least negatively charged in the *in-vacuo* state (models **G**). The same trend holds for the positive charges of the hydrogen atoms. For both oxygen and hydrogen atoms, the crystal environment **C** produces the atomic charges that are most similar to the protein environment. All these trends are reflected in the interaction ESP plots in the same way (Figure 5). For the carbon atoms, which are not involved in the hydrogen bonding network, the same trend is true, but less pronounced: the positive charges are highest in the **P** model and lowest in the **G** models.

On average, the charges produced in the Hirshfeld cluster-charge model (**CC**) are slightly higher than those in the QM/MM crystal model **C**, sometimes even higher than in the enzyme model **P** – an overestimation of the polarization already discussed with respect to Figure 5 h,i). However, the atomic charges produced after XCW fitting are significantly higher than the charges of any of the other models. Despite these discrepancies, chemically meaningful charge differences of the same element in different functional groups are always preserved. This refers to the lower negative charge of the epoxide oxygen atom O1 or the higher positive charges of the carboxyl and carbonyl carbon atoms C3 and C4 compared to the epoxide carbon atoms C1 and C2.

Table 2. QTAIM atomic charges (e) of atoms in each model. For atom labeling, see Figure 1a. For model definition, see Figure 10. The isolated-anion and the QM/MM models do not include explicit K cations, but there are two symmetry-independent ones in the crystal structure (CC and XCW models).

Atom	G (w/o K)	G (w K)	S (w/o K)	S (w K)	C	P	CC	XCW
O1	-0.875	-0.852	-0.903	-0.898	-0.899	-0.907	-0.928	-1.140
O2	-1.212	-1.227	-1.257	-1.247	-1.267	-1.303	-1.346	-1.266
O3	-1.216	-1.230	-1.256	-1.245	-1.258	-1.304	-1.276	-1.389
O4	-1.151	-1.133	-1.193	-1.192	-1.222	-1.277	-1.225	-1.276
N1	-1.105	-1.126	-1.153	-1.152	-1.201	-1.165	-1.185	-1.563
C1	0.356	0.359	0.356	0.357	0.360	0.382	0.398	0.325
C2	0.335	0.364	0.341	0.346	0.349	0.390	0.377	0.637
C3	1.668	1.581	1.627	1.611	1.614	1.650	1.724	1.820
C4	1.350	1.404	1.409	1.412	1.393	1.407	1.500	1.583
H1	0.052	0.043	0.071	0.074	0.068	0.094	0.053	-0.010
H2	0.029	0.072	0.066	0.072	0.063	0.084	0.053	0.150
H3	0.387	0.396	0.445	0.446	0.506	0.495	0.457	0.647
H4	0.384	0.409	0.448	0.449	0.495	0.456	0.435	0.569
K	-	0.940	-	0.968	-	-	0.960/0.971	0.925/0.957

Conclusions

In this study we used model compound **1** which consists of the pure deprotonated epoxysuccinyl peptide group that is the biologically active center of the drug loxistatin acid (E64c). In molecular dynamics simulations, **1** showed to exhibit the same stable intermolecular contacts inside the enzyme pocket of cathepsin-B that lead to the irreversible inhibition of this protease by E64c. Moreover, in the crystal structure **1**K·H_2O intermolecular interactions (hydrogen bonds and electrostatic contacts) are formed that are very similar to those in the enzyme pocket, involving biologically important water molecules and potassium cations. The K^+ ion is located in the crystal where the His^+ group is located in cathepsin B.

Detailed analyses of the interaction densities and electrostatic potentials as well as of the difference electron densities show unambiguously that the crystal environment, as described by QM/MM, is significantly more similar to the protein environment than any of the other environments (a cluster of Hirshfeld point charges and dipoles, a solvation model, or the isolated molecule in vacuum, the latter two with or without explicit perturbation by the K^+ counter-cation). The difference between the integrated interaction densities in the protein and crystal environments was measured to be smaller than 0.1e over the entire anion **1**, or 0.075 units of the RSR value when integrated over all bonds of **1**. We conclude that the sentiment often expressed in the literature that the interaction density in crystals of the pure substance is a measure of the polarization of the same molecule in the enzyme active site is true for the analyzed compound. We also showed that the interaction density can be used to predict the features of the electrostatic recognition process.

Finally, we investigated whether the above conclusion means that experimental electron-density studies are an ideal tool to get deeper insight into the enzyme-ligand interaction. Certainly, with an error-free experiment and an ideal model to refine the experimental structure factors, such studies would be a close simulation of the biological situation. However, the crystal field effect on the electron density is relatively small, i.e. less than one out of 67 electrons of **1** are shifted due to polarization. Therefore, the interpretation of the experimental results and maps is difficult relative to the idealized theoretical models. The accuracy of the data and the (in-)flexibility of the model have a huge influence on the result. The convolution of many physical effects in the experimental data that might be lacking in the theoretical ansatz/model does not allow to unambiguously pin down the crystal field effect in the difference densities. Certainly, for an effect such as the polarization in different environments we cannot trust the experimental data blindly but have to critically assess them, when, simultaneously, a computationally relatively cheap and simple method such as a COSMO solvation model already approximates the results quite closely.

Experimental and Computational Section

The synthesis and spectroscopic characterization of **1**K·H_2O has been reported previously.[19] Details of the synchrotron X-ray diffraction experiment are given in Table 3. A high-resolution measurement was carried out at beamline BL02B1 of SPring-8 using a helium gas-stream for cooling down to 25 K and a large curved image plate for detection. The data resolution was limited to d = 0.45 Å after radiation-damage problems were observed, which was corrected by using a scaling procedure in the software *RapidAuto*[44] (see more details in the Supporting Information). The crystal structure of **1**K·H_2O was solved using *ShelxT* and refined using *olex2.refine* and *Tonto* within the *NoSpherA2* framework.[45-47] HAR was performed on a level of theory of B3LYP/def2-TZVP using a radius of cluster charges and dipoles of 20 Å. Anisotropic displacement parameters of hydrogen atoms of **1** were determined using the *SHADE2.1* webservice,[48] and then iteratively updated in *Shade* in-between HAR cycles, whereas the hydrogen atoms of the water molecule were refined with isotropic displacement parameters.

The crystal structure of **1**K·H_2O, the conditions and the parameters used in the HAR as well as the structure factor magnitudes are deposited with the Cambridge Structural Database with the reference code CCDC-2024395. The final wavefunction calculation after the last structural least-squares refinement in HAR serves as the model **CC** since it is a single-point calculation under the influence of a surrounding cluster of self-consistent Hirshfeld point charges and dipoles. Subsequent XCW fitting was performed using a level of theory of B3LYP/def2-TZVP without cluster charges to include an estimate of the electron correlation effect into the wavefunction ansatz whereas the crystal environmental effect is fitted based on the experimental data. The resulting wavefunction serves as the model **X** (or XWR/XCW).

The final crystal structure was used to build a supercluster according to the symmetry of the crystal to perform QM/MM calculations in the crystal system (**C**), treating one anion **1** in the center of the cluster quantum-mechanically and the proximity using molecular mechanics based on a *CHARMM* force field. Parameters describing **1** were obtained from the swissparam webservice.[49] QM/MM structure optimization were performed using *NAMD2* and *Turbomole* in all cases during this study, interfaced using *ChemShell*.[50-54] The level of theory used for the QM part was B3LYP/def-TZVP. Previous studies were able to show that the density is sufficiently well described using triple zeta basis sets, and the extent of differences using different levels of theory in QM/MM are documented.[38-40]

Table 3. Crystal, measurement, and refinement details for **1**K·H_2O.

	HAR	**XCW**
Empirical formula	$C_4H_6KNO_5$	
Formula weight (*g mol⁻¹*)	187.20	
Cell setting, space group	Trigonal, $P3_12$ (No. 152)	
Z	6	
Temperature (*K*)	25	
Unit cell dimensions		
a (*Å*)	8.14360 (10)	
c (*Å*)	19.1623 (3)	
V (*Å³*)	1100.55 (3)	
F(000)	576	
Crystal size (mm)	0.15 x 0.11 x 0.10	
Crystal form, color	Block, colorless	
Wavelength λ (*Å*)	0.3532	
Absorption correction	Multi-scan	
Absorption coefficient μ (*mm⁻¹*)	0.115	
T_{min}/T_{max}	0.884/1.206	
Max. θ (°)	23.103	
$(\sin\theta/\lambda)_{max}$ (*Å⁻¹*)	1.11094	
Measured, independent and observed reflections	62635, 8415, 8362	
Completeness	100%	
Redundancy	7.474	
Weighting scheme [a] w=	$(\sigma^2+(0.048P)^2+0.162P)^{-1}$	σ^{-2}
$R_{int}(F^2)$	0.0327	- (merged data)
Number of parameters	121	121
$R_1(F)$	0.0266	0.0200
$R_w(F)$	0.0757	0.0208
$R_{all}(F)$	0.0267	-[b]
N_{ref}/N_v	69.1	
$\Delta\rho_{max}$, $\Delta\rho_{min}$ (*e Å⁻³*)	1.371/-0.378	0.756/-0.216
Flack parameter	0.13 (6)	-[b]
	[a] with $P=\frac{F_O^2+2F_C^2}{3}$	[b] not reported by tonto
CCDC deposition no.	2024395	

For a model of the protein environment, **1** was placed into the active site of cathepsin B, based on a protein crystal structure of the E64c – cathepsin B complex[17] already modified as MD input including all hydrogen atoms in ref. [13a]. The resulting geometry was then equilibrated and simulated with molecular dynamics (MD) at 310 K using a *CHARMM* force field for protein, water and salt ions.[55-57] The geometries at 4 points during the simulation between 4 and 5 ns (250 ps apart) were taken to perform QM/MM geometry optimizations (*NAMD, Turbomole, ChemShell*), optimizing the geometry of **1** quantum-mechanically on a level of theory of B3LYP/def-TZVP, and optimizing the residues in proximity of **1** with at least one atom in a range of 3 Å with molecular mechanics. Since 4 slightly different geometries were the result of these calculations, the calculation of grids was performed on all wavefunctions and the arithmetic average of the geometries and the grid values were calculated for the comparisons between the different environments. All plots labelled as model **P** refer to these averaged grids.

Optimized geometries in solvation (**S**) and in-vacuo (commonly called gas-phase model, **G**) were obtained by ab-initio optimizations using the level of theory B3LYP/def-TZVP in *Turbomole*.[52] For model **S**, a *COSMO* solvation model of water was used. These calculations were carried out with anionic molecule **1**, as well as with a potassium counter-cation, to give an interaction partner as a first model of interaction with the environment.

The calculation of grid files of the interaction density and interaction ESP is simple, however, the setup of the boxes of the grid files for bond-wise comparison between different geometries is difficult. We programmed an in-house software (*cuQCT*, author: Florian Kleemiss) for the calculation and manipulation of grid files in the Gaussian cube format. Bond-centered grid files are generated by first finding the midpoint between the two positions of selected atoms. The unit vector connecting them will provide the vector $\vec{a}$ of the box. A third atom is used to form a plane and to find the vector in this plane that is perpendicular to $\vec{a}$. This way, the second vector for the grid is found, which will be called $\vec{b}$. Both vectors are normalized to have a length of 1 bohr. The last vector $\vec{c}$ for the calculation of the grid is found by the calculation of the vector product of $\vec{a}$ and $\vec{b}$. The midpoint of the first two selected atoms will then be used as the central point of the grid. By choosing a multiplicator for the distance between atoms 1 and 2 in all three dimensions of the grid, the atoms in different grid files will always be located at the same position of the grids. The origin of the grid is then found by moving into the negative directions of $\vec{a}$, $\vec{b}$ and $\vec{c}$ for half the corresponding length of the complete grid size from the central point. Then the calculation with a fixed number of gridpoints scales the grid vectors and allows comparability with grids in different settings or with different molecular geometry without interpolations.

To have a quantitative measure of similarity between two ED or ESP grids, a real space R value (RSR or R_{RS}) was used. The definition of the RSR was taken from the literature.[58] For the ED, it is defined as:

$$R_{RS}=\frac{\sum|\rho_1(r)-\rho_2(r)|}{\sum|\rho_1(r)+\rho_2(r)|}.$$

This RSR gives insight in the total difference in ED relative to the total ED of the compound. It is likewise defined for the ESP. A value of 1 corresponds to a total shift of the complete property, while 0 corresponds to perfect agreement. As an alternative, the weighted Jaccord or Soergel distance[59], as a variation of the Jaccord distance for binary groups[60], can be defined for the ED differences as:

$$d_{Jaccord}=\frac{\sum\min\big(\rho_1(r),\rho_2(r)\big)}{\sum\max\big(\rho_1(r),\rho_2(r)\big)},$$

where the sum runs over all grid points of the density.

The number of electrons shifted between two ED distributions is given by the sum over the whole grid with grid point indices i, j, k. The local density difference is multiplied with the voxel size (which is the triple product of the grid vectors) and divided by 2 since local depletion of electrons will accumulate in a different place:

$$N_e=\sum_i\sum_j\sum_k\left[\frac{\rho_1(r)-\rho_2(r)}{2}*\vec{a}\cdot\left(\vec{b}\times\vec{c}\right)\right]$$

The wavefunctions for the models **CC** and **X** of the crystal system required the explicit inclusion of two potassium ions and a water molecule in the asymmetric units of the crystal structure dictated by the crystallographic symmetry. This is necessary to describe the cluster of charges correctly in the case of Hirshfeld point charges and to calculate correct structure factors in the case of XCW. However, for an optimum comparability with respect to the QM/MM calculations, the calculation of ESP and ED grid files for the models **CC** and **X** required a routine to calculate these grids without explicit, just implicit contributions of the potassium ions and atoms in water. In a new feature of *cuQCT*, these atoms were ignored for the calculation of ESP and ED by skipping all basis functions that are associated to these atoms while calculating the value of a molecular orbital at a point in space for the ED or during the integration of the density for the ESP.

Acknowledgement

S. Grabowsky acknowledges funding of the German Research Foundation (DFG) via the Emmy-Noether project GR 4451/1-1 and of the Australian Research Council (ARC) via Discovery Project DP110105347. F. Kleemiss thanks the German Academic Scholarship Foundation (Studienstiftung des Deutschen Volkes) for a scholarship. The synchrotron measurement at SPring-8 was carried out under proposal number 2013B1056.

References

[1] (a) H.-J. Böhm, G. Klebe, *Angew. Chem. Int. Ed. Engl.* **1996**, *35*, 2588−2614; b) R. E. Babine, S. L. Bender, *Chem. Rev.* **1997**, *97*, 1359−1472.

[2] D. E. Koshland, *Angew. Chem. Int. Ed.* **1995**, *33*, 2375−2378.

[3] a) S. Sarkhel, G. R. Desiraju, *Proteins: Struct. Funct. Bioinf.* **2004**, *54*, 247–259; b) A. R. Fersht, *Trends Biochem. Sci.* **1987**, *12*, 301−304.

[4] M. W. Shi, A. N. Sobolev, T. Schirmeister, B. Engels, T. C. Schmidt, P. Luger, S. Mebs, B. Dittrich, Y.-S. Chen, J. M. Bąk, D. Jayatilaka, C. S. Bond, M. J. Turner, S. G. Stewart, M. A. Spackman, S. Grabowsky, *New J. Chem.* **2015**, *39*, 1628−1633.

[5] C. Pascard, *Acta Cryst. D* **1995**, *51*, 407−417.

[6] a) H. F. Velec, H. Gohlke, G. Klebe, *J. Med. Chem.* **2005**, *48*, 6296–6303; b) I. J. Bruno, J. C. Cole, P. M. Jos, R. S. Lommerse, R. Taylor, M. L. J. Verdonk, *Computer-Aided Mol. Des.* **1997**, *11*, 525–537.

[7] a) G. Klebe, *J. Mol. Biol.* **1994**, *237*, 212−235; b) M. Böhm, G, Klebe, *J. Med. Chem.* **2002**, *45*, 1585−1597.

[8] a) M. L. Verdonk, J. C. Cole, P. Watson, V. Gillet, P. Willett, *J. Mol. Biol.* **2001**, *307*, 841−859; b) D. R. Boer, J. Kroon, J. C. Cole, B. Smith, M. L. Verdonk, *J. Mol. Biol.* **2001**, *312*, 275−287.

[9] a) G. Náray-Szabó, *J. Mol. Graphics* **1989**, *7*, 76−81; b) E. Kangas, B. Tidor, *J. Phys. Chem. B* **2001**, *105*, 880−888; c) N. Muzet, B. Guillot, C. Jelsch, E. Howard, C. Lecomte, *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 8742–8747; d) S. Mebs, A. Lüth, P. Luger, *Bioorg. Med. Chem.* **2010**, *18*, 5965–5974.

[10] a) S. Grabowsky, D. Jayatilaka, R. F. Fink, T. Schirmeister, B. Engels, *Z. Anorg. Allg. Chem.* **2013**, *639*, 1905–1921; b) T. Leduc, E. Aubert, E. Espinosa, C. Jelsch, C. Iordache, B. Guíllot, *J. Phys. Chem. A*, **2019**, *123*, 7156–7170.

[11] P. Luger, *Org. Biomol. Chem.* **2007**, *5*, 2529–2540.

[12] a) D. E. Hibbs, C. J. Austin-Woods, J. A. Platts, J. Overgaard, P. Turner, *Chem. Eur. J.* **2003**, *9*, 1075–1084; b) S. Grabowsky, T. Pfeuffer, L. Chęcińska, M. Weber, W. Morgenroth, P. Luger, T. Schirmeister, *Eur. J. Org. Chem.* **2007**, 2759–2768; c) N. E. Ghermani, A. Spasojević-de Biré, N. Bouhmaida, S. Ouharzoune, J. Bouligand, A. Layre, R. Gref, P. Couvreur, *Pharm. Res.* **2004**, *21*, 598–607; d) A. Wagner, R. Flaig, B. Dittrich, H. Schmidt, T. S. Koritsánszky, P. Luger, *Chem. Eur. J.* **2004**, *10*, 2977–2982; e) B. Dittrich, T. S. Koritsánszky, A. Volkov, S. Mebs, P. Luger, *Angew. Chem. Int. Ed.* **2007**, *46*, 2935–2938; f) D. Parrish, E. A. Zhurova, K. Kirschbaum, A. A. Pinkerton, *J. Phys. Chem. B* **2006**, *110*, 26442–26447; g) E. A. Zhurova, C. F. Matta, N. Wu, V. V. Zhurov, A. A. Pinkerton, *J. Am. Chem. Soc.* **2006**, *128*, 8849–8861; h) E. J. Yearly, E. A. Zhurova, V. V. Zhurov, A. A. Pinkerton, *J. Am. Chem. Soc.* **2007**, *129*, 15013–15021; i) R. Destro, R. Soave, M. Barzaghi, L. Lo Presti, *Chem. Eur. J.* **2007**, *13*, 6942–6956; j) J. Overgaard, I. Turel, D. E. Hibbs, *Dalton Trans.* **2007**, 2171–2178; k) D. E. Hibbs, J. Overgaard, S. T. Howard, T. H. Nguyen, *Org. Biomol. Chem.* **2005**, *3*, 441–447.

[13] a) M. Mladenovic, M. Arnone, R. F. Fink, B. Engels, *J. Phys. Chem. B* **2009**, *113*, 5072-5082; b) B. Engels, T. C. Schmidt, C. Gatti, T. Schirmeister, R. F. Fink, in *Electron Density and Chemical Bonding II*, Springer, **2011**, pp. 47-97.

[14] K. Hanada, M. Tamai, M. Yamagishi, S. Ohmura, J. Sawada, I. Tanaka, Agric. Biol. Chem. **1978**, 42, 523−528.

[15] a) H.-H. Otto, T. Schirmeister, *Chem. Rev.* **1997**, *97*, 133−171; b) J. C. Powers, J. L. Asgian, Ö. D. Ekici, K. E. James, *Chem. Rev.* **2002**, *102*, 4639−4750.

[16] a) M. Mladenovic, K. Junold, R. F. Fink, W. Thiel, T. Schirmeister, B. Engels, *J. Phys. Chem. B.* **2008**, *112*, 5458−5469; b) M. Mladenovic, K. Ansorg, R. F. Fink, W. Thiel, T. Schirmeister, B. Engels, *J. Phys. Chem. B* **2008**,*112*,11798–11808.

[17] A. Yamamoto, K. Tomoo, K. Matsugi, T. Hara, Y. In, M. Murata, K. Kitamura, T. Ishida, *Biochim. Biophys. Acta* **2002**, *1597*, 244–251.

[18] a) S. Grabowsky, T. Pfeuffer, W. Morgenroth, C. Paulmann, T. Schirmeister, P. Luger, *Org. Biomol. Chem.* **2008**, *6*, 2295–2307; b) S. Grabowsky, T. Schirmeister, C. Paulmann, T. Pfeuffer, P. Luger, *J. Org. Chem.* **2011**, *76*, 1305–1318.

[19] M. W. Shi, S. G. Stewart, A. N. Sobolev, B. Dittrich, T. Schirmeister, P. Luger, M. Hesse, Y.-S. Chen, P. R. Spackman, M. A. Spackman, S. Grabowsky, *J. Phys. Org. Chem.* **2017**, *30*, e3683.

[20] a) S. Grabowsky, P. Luger, J. Buschmann, T. Schneider, T. Schirmeister, A. N. Sobolev, D. Jayatilaka, *Angew. Chem. Int. Ed.* **2012**, *51*, 6776–6779; b) L. Chęcińska, W. Morgenroth, C. Paulmann, D. Jayatilaka, B. Dittrich, *Cryst. Eng. Comm.* **2013**, *15*, 2084–2090.

[21] a) D. Jayatilaka, B. Dittrich, *Acta Cryst. A* **2008**, *64*, 383–393; b) S. C. Capelli, H.-B. Bürgi, B. Dittrich, S. Grabowsky, D. Jayatilaka, *IUCrJ* **2014**, *1*, 361–379; c) M. Woińska, S. Grabowsky, P. M. Dominiak, K. Woźniak, D. Jayatilaka, Sci. *Adv.* **2016**, *2*, e1600192.

[22] a) D. Jayatilaka, *Phys. Rev. Lett.* **1998**, *80*, 798–801. b) D. Jayatilaka, D. J. Grimwood, *Acta Cryst. A* **2001**, *57*, 76–86. c) D. J. Grimwood, D. Jayatilaka, *Acta Cryst. A* **2001**, *57*, 87–100.

[23] S. Grabowsky, A. Genoni, H.-B. Bürgi, *Chem. Sci.* **2017**, *8*, 4159–4176.

[24] B. Dittrich, C. F. Matta, *IUCrJ* **2014**, *1*, 457–469.

[25] a) M. Krijn, H. Graafsma, D. Feil, *Acta Cryst. B* **1988**, *44*, 609–616; b) M. A. Spackman, P. G. Byrom, M. Alfredsson, K. Hermansson, *Acta Cryst. A* **1999**, *55*, 30–47; c) R. Y. de Vries, D. Feil, V. G. Tsirelson, *Acta Cryst. B* **2000**, *56*, 118–123; d) B. Dittrich, M. A. Spackman, *Acta Cryst. A* **2007**, *63*, 426–436.

[26] P. Coppens, A. Volkov, *Acta Cryst. A* **2004**, *60*, 357–364.

[27] a) R. Dovesi, M. Causa', R. Orlando, C. Roetti, V. R. Saunders, *J. Chem. Phys.* **1990**, *92*, 7402–7411; b) D. Feil, *J. Mol. Struct.* **1990**, *237*, 33–46; c) B. Dittrich, E. Sze, J. J. Holstein, C. B. Hübschle, D. Jayatilaka, *Acta Cryst. A* **2012**, *68*, 435–442.

[28] M. Ernst, A. Genoni, P. Macchi, *J. Mol. Struct.* **2020**, *1209*, 127975.

[29] a) E. Espinosa, C. Lecomte, N. E. Ghermani, J. Devémy, M. M. Rohmer, M. Bénard, E. Molins, *J. Am. Chem. Soc.* **1996**, *118*, 2501–2502; b) I. Mata, E. Molins, I. Alkorta, E. Espinosa, *J. Phys. Chem. A* **2007**,*111*, 6425–6433; c) C. F. Matta, *J. Comput. Chem.* **2014**, *35*, 1165–1198.

[30] F. Javier Luque, M. Orozco, *J. Comput. Chem.* **1998**, *19*, 866–881.

[31] A. Madsen, *J. Appl. Cryst.* **2006**, *39*, 757–758.

[32] O. V. Dolomanov, L. J. Bourhis, R. J. Gildea, J. A. K. Howard, H. Puschmann, *J. Appl. Cryst.* **2009**, *42*, 339–341.

[33] L. Bučinský, D. Jayatilaka, S. Grabowsky, *Acta Cryst. A* **2019**, *75*, 705–717.

[34] A. Genoni, L. H. R. Dos Santos, B. Meyer, P. Macchi, *IUCrJ* **2017**, *4*, 136–146.

[35] N. K. Hansen, P. Coppens, *Acta Cryst. A* **1978**, *34*, 909–921.

[36] W. Humphrey, A. Dalke, K. Schulten, *J. Mol. Graphics* **1996**, *14*, 33–38.

[37] H. Wolf, M. R. V. Jørgensen, Y.-S. Chen, R. Herbst-Irmer, D. Stalke, *Acta Cryst. B* **2015**, *71*, 10–19.

[38] N. Kocher, J. Henn, B. Gostevskii, D. Kost, I. Kalikhman, B. Engels, D. Stalke, *J. Am. Chem. Soc.*, **2004**, *126*, 5563-5568.

[39] D. Leusser, J. Henn, N. Kocher, B. Engels, D. Stalke, *J. Am. Chem. Soc.*, **2004**, *126*, 1781-1793.

[40] J. Henn, D. Ilge, D. Leusser, D. Stalke, B. Engels, *J. Phys. Chem.*, **2004**, *108*, 9442-9452.

[41] T. Lu, F. Chen, *J. Comp. Chem*. **2012**, *33*, 580–592.

[42] M. Woińska, D. Jayatilaka, B. Dittrich, R. Flaig, P. Luger, K. Woźniak, P. M. Dominiak, S. Grabowsky, *Chem. Phys. Chem.* **2017**, *18*, 3334–3351.

[43] R. F. W. Bader, *Atoms in Molecules - A Quantum Theory*; Oxford University Press: Oxford, U.K., **1990**.

[44] Rigaku (2004). RAPID-AUTO. Rigaku Corporation, Tokyo, Japan.

[45] G. M. Sheldrick, Acta Cryst. A **2015**, *71*, 3-8.

[46] D. Jayatilaka and D. J. Grimwood, *Computational Sciece* - ICCS 2003, 142-151. (https://github.com/dylan-jayatilaka/tonto.git, commit d26b49d)

[47] F. Kleemiss, O. V. Dolomanov, M. Bodensteiner, N. Peyerimhoff, L. Midgley, L. J. Bourhis, A. Genoni, L. A. Malaspina, D. Jayatilaka, J. L. Spencer, F. White, B. Grundkötter-Stock, S. Steinhauer, D. Lentz, H. Puschmann, S. Grabowsky, submitted to Chem. Sci., October 2020.

[48] A. Ø. Madsen, J. Appl. Cryst. **2006**, *39*, 757-758.

[49] Zoete, M. A. Cuendet, A. Grosdidier, O. Michielin, *J. Comput. Chem.* **2011**, *32*, 2359-2368.

[50] L. Kalé, R. Skeel, M. Bhandarkar, R. Brunner, A. Gursoy, N. Krawetz, J. Phillips, A. Shinozaki, K. Varadarajan, K. Schulten, *J. Comput. Phys.* **1999**, *151*, 283-312.

[51] J. C. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R. D. Skeel, L. Kale, K. Schulten, *J. Comput. Chem.* **2005**, *26*, 1781–1802.

[52] R. Ahlrichs, M. Baer, M. Haeser, H. Horn, C. Koelmel, *Chem. Phys. Lett.* **1989**, *162*, 165–169. (Used version of Turbomole: 7.0.1)

[53] P. Sherwood, A. H. de Vries, M. F. Guest, G. Schreckenbach, C. R. A. Catlow, S. A. French, A. A. Sokol, S. T. Bromley, W. Thiel, A. J. Turner, S. Billeter, F. Terstegen, S. Thiel, J. Kendrick, S. C. Rogers, J. Casci, M. Watson, F. King, E. Karlsen, M. Sjøvoll, A. Fahmi, A. Schäfer, Ch. Lennartz, *J. Mol. Struct. (Theochem.)* **2003**, *632*, 1-28.

[54] J. Kästner, J. M. Carr, T. W. Keal, W. Thiel, A. Wander, P. Sherwood, *J. Phys. Chem. A*, **2009**, *113*, 11856.

[55] A. D. MacKerell, M. Feig, C. L. Brooks, *J. Comput. Chem.* **2004**, *25*, 1400-1415.

[56] A. D. MacKerell Jr, D. Bashford, M. Bellott, R. L. Dunbrack Jr, J. D. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, *J. Phys. Chem. B* **1998**, *102*, 3586-3616.

[57] R. B. Best, X. Zhu, J. Shim, P. E. M. Lopes, J. Mittal, M. Feig, A. D. MacKerell, *J. Chem. Theory Comput.* **2012**, *8*, 3257–3273.

[58] T. A. Jones, J.-Y. Zou, S. T. Cowan, M. Kjeldgaard, *Acta Cryst. A* **1991**, *47*, 110-119.

[59] J. C. Gower, Similarity, Dissimilarity, and Distance Measure. In *Encyclopedia of Biostatistics*, **2005**.

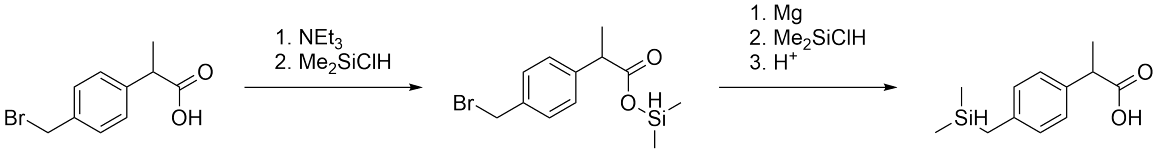
[60] P. Jaccard, *New Phytologist*, **1912**, *11*, 37-50.

5.2 Sila-Ibuprofen

It was shown in chapter 4 that the procedure of XWR can reproduce effects missed by a selected wavefunction *ansatz* and already medium resolution of the diffraction data is sufficient given the quality of the data is good. Also, in section 5.1, it was shown that the similarity between the crystal influence on the molecular electron density is highest compared to the situation in an enzyme. The effect of polarization had a significant influence on the observed electrostatic potential and electron density, especially around hydrogen atoms. The question arises, whether a change in the polarization of a molecule itself through an exchange of atoms will change its behaviour, since the polarization through the environment might be affected as well. If the amount of polarization by the environment is changed it might also have influences on the energetic situation of, for example, the binding of a drug in the enzyme. Since carbon-silicon-exchange offers a tool to change electrostatic properties of a molecule without changing the functional group. A carbon-silicon exchanged *tert*-butyl functional group in principle has the same shape and angles between bonds, except for changed electrostatic properties. This might allow conceptual creation of a drug-analogous with changed electrostatic properties while retaining its functionality, rendering them bioisosters with different polarizabilities and polarization. [240, 291, 292] The following project applies this concept to the essential drug ibuprofen.
Although major parts of this projects are mentioned in the publication "*Sila-Ibuprofen*" (see section 5.2.14) and the supporting information of this publication (see appendix D.2) the results and insight won by this projects and the method development carried out for it are too manifold to be presented in the form of a communication type of article. Therefore some of the results will be discussed in a bigger context in the following sections and paragraphs.

5.2.1 Synthesis

The synthesis of carbon-silicon-exchanged ibuprofen (compare 1.7), called sila-ibuprofen, is achieved by using commercially available 2-[(4-bromomethyl)phenyl]propionic acid, dissolved in diethylether and protected by a dimethylsilyl group, through addition of 4 equivalents of dimethylchlorosilane and 2 equivalents of triethylamine under inert conditions. After filtration of the precipitated triethylammonium chloride, the mixture was added drop wise to an excess of activated magnesium turnings in diethylether under reflux. The mixture was refluxed for at least 10 hours and then ice water was added. The organic phase was washed again three times with distilled cold water and the solvent was removed. The yielding sticky oil was purified by column chromatography by flushing three column volumes of n-hexane and subsequently eluting the product using ethyl acetate. The product was obtained in 85 % yield after removal of the solvent.
The first step of the synthesis is the formation of the dimethylsilyl ester of the 2-[(4-bromomethyl)phenyl]propionic acid. The triethylamine acts as a base and deprotonates the acid, which can then undergo a nucleophilic reaction with the chlorosilane forming the silane-ester and triethylammonium chloride, which precipitates in the organic medium as a colourless salt. This ester is formed in order to deactivate the acid function for the consecutive reaction. A *Barbier* reaction, where magnesium turnings are activated and addition of the reaction mixture after filtration results in the raw product forms the protected version of the product molecule. After work up and cleaning, using water for extraction of unwanted reaction products like the magnesium bromide chloride and other possible side products and column chromatography the final product was obtained in 85 % yield. This simple setup which is basically a one-pot reaction allows easy realization on a bigger scale and without complex apparatus, e.g. on industrial scales. The reaction scheme is visualized in Figure 5.1.
Using this material, full chemical characterization was performed with the help of the group of Prof. Beckmann in Bremen. The first time this synthesis was conducted by Aileen Justies and Nelly R. King in Prof. Beckmann's group at the Free Univserity of Berlin. The synthesis was carried out for further analysis and crystallizations by Pim Puylaert, Daniel Duvinage and me.



3

2

Figure 5.1: Scheme of synthesis of sila-ibuprofen using a *Barbier* reaction.

5.2.2 Characterization

A full chemical characterization of sila-ibuprofen was performed in order to understand the properties and possible advantages the sila-substitution might have imposed. The collaboration of the groups of Prof. Dringen, Prof. Beckmann, Prof. Grabowsky and other institutes in the department of chemistry in Bremen allowed a broad variety of methods to be used, which will be presented in the following subsections.
In the following sections the melting enthalpy, solubility, stability in aqueous media, NMR spectra and full crystallographic analysis will be shown.

5.2.2.1 Melting Enthalpy

Since the melting enthalpy and melting point were major reasons for the concept of sila-substitution in the first place, these were measured using Differential Scanning Calorimetry (DSC) by Prof. Anne Staubitz. In fact the melting point of sila-ibuprofen (45-47.4 °C) is significantly lower compared to conventional ibuprofen, which has a melting point around 74 °C. [463–465] Also the melting enthalpy measured as 15.8 ± 0.5 kJ mol^{-1} reflects this trend, as it is ca. 10 kJ mol^{-1} lower than that of ibuprofen (26.7 kJ mol^{-1} [466]). This data suggests a lower lattice energy of sila-ibuprofen, as the lower amount of energy needed for melting indirectly indicates lower lattice energy. If the lattice energy is lower a better solubility might be one of the resulting properties of the material.
The melting enthalpy of both, ibuprofen and sila-ibuprofen, was determined by DSC, using a Metler-Toledo DSC3+ instrument, where 40 μL aluminium crucibles, a pin and a pierced lid were used to measure in reference to an empty crucible. The temperature program of ibuprofen ranged from 25°C to 125°C, the program for sila-ibuprofen started at 0°C and ranged to 70°C where both programs had a temperature ramping rate of 10 K min^{-1} and a N_2 flow of 20 mL min^{-1}. These measurements were performed by Prof. Dr. Anne Staubitz and data evaluation performed by her using the software *Star-e* Version 15.01.

5.2.2.2 Solubility

One of the limiting factors of the application of classical ibuprofen is its limited solubility of 21 mg L^{-1} in water. [302] The pure substance of sila-ibuprofen exhibits a low melting point, which might be seen as an indicator to lowered lattice energies, which might in turn yields better solubility in aqueous media. This would be highly desired for medicinal application purposes in intra-venous application or the application in wounds.
High performance liquid chromatography-ultra violet detection (HPLC-UV) is a common technique used to determine concentrations of a drug molecule in solution. This technique was also employed for the determination of the solubility of sila-ibuprofen. It was performed using a RP-18 gravity column in a HPLC-UV experiment detecting at a wavelength of 235 nm. This was done using ibuprofen as an internal standard with a fixed concentration of 0.1 mg L^{-1} and equidistant calibration from 5 mg L^{-1} to 70 mg L^{-1} of sila-ibuprofen prepared from a stock solution of sila-ibuprofen in ethanol. An eluent of 1:1 water:acetonitrile with 0.1 mol L^{-1} acetic acid in was used for the calibration in triple determination. The final calibration plot is shown in Figure 5.2.
The calibration shows linear behaviour and allows easy determination of the concentration of a test solution, where sila-ibuprofen was added to distilled water and stirred for two days in order to reach saturation of the solution. Then this test solution was treated identically to the calibration solutions and measured in triple determination. The solubility in water of pure sila-ibuprofen was determined to be 83 ± 3 mg L^{-1}. This

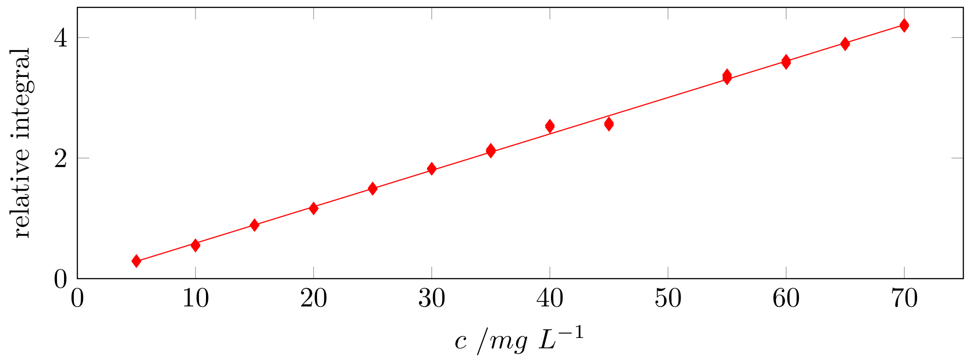


Figure 5.2: Relative integrals between the integral of the absorbance peak of sila-ibuprofen and the internal standard ibuprofen against the concentration of sila-ibuprofen with linear fit of data-points (R^2 = 0.998). Points at 50 mg L^{-1} were discarded due to machine failure.

is almost four times higher than the solubility of conventional ibuprofen, which is documented to be 21 mg L^{-1}. [302]

5.2.2.3 NMR Spectra

All Nuclear Magnetic Resonance (NMR) spectra mentioned in this work were performed on a Bruker Avance 600 spectrometer and are reported on the δ(ppm) scale and referenced against $SiMe_4$. ^1H, $^{13}C\{^1H\}$ and $^{29}Si\{^1H\}$ chemical shifts are reported relative to the residual peak of the solvent ($(CD_3)(CD_2H)CO$ at 2.09 ppm for $(CD_3)_2CO$) in the ^1H spectra and to the peak of the deuterated solvent $(CD_3)_2CO$ at 30.60 ppm in the $^{13}C\{^1H\}$ NMR spectra, while $^{29}Si\{^1H\}$ spectra are referenced against $SiMe_4$. The spectra were collected with help of Daniel Duvinage and Dr. Pim Puylaert.

The spectra of sila-ibuprofen after chromatography column purification are shown in Figures 5.3 to 5.5.

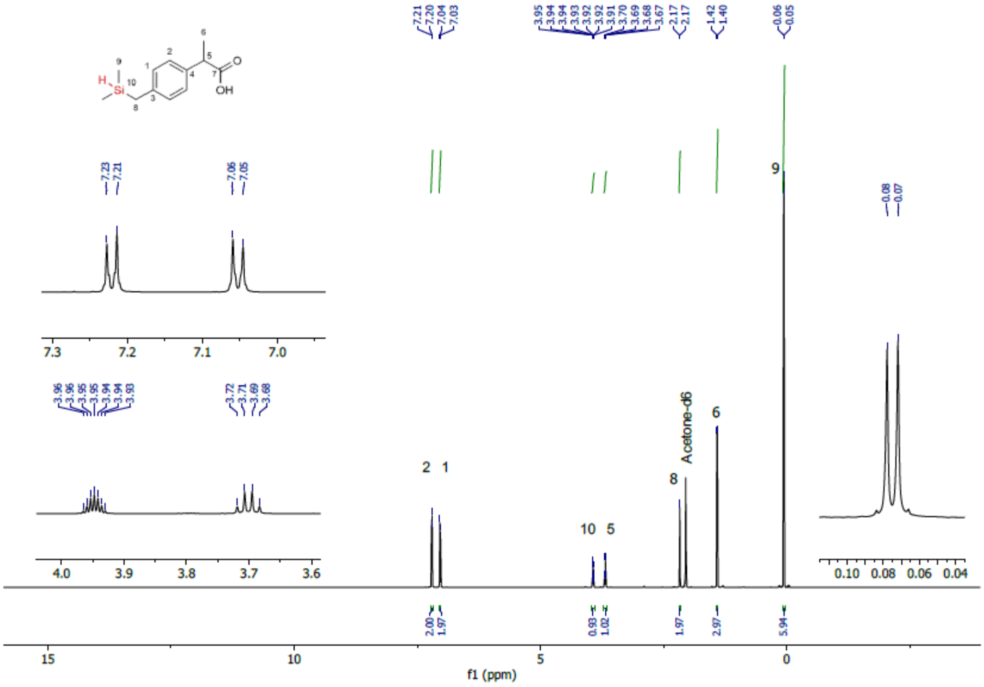


Figure 5.3: ^1H-NMR of sila-ibuprofen measured in acetone-d6 at 600MHz. Signals are assigned numbers according to labeling scheme in top left corner.

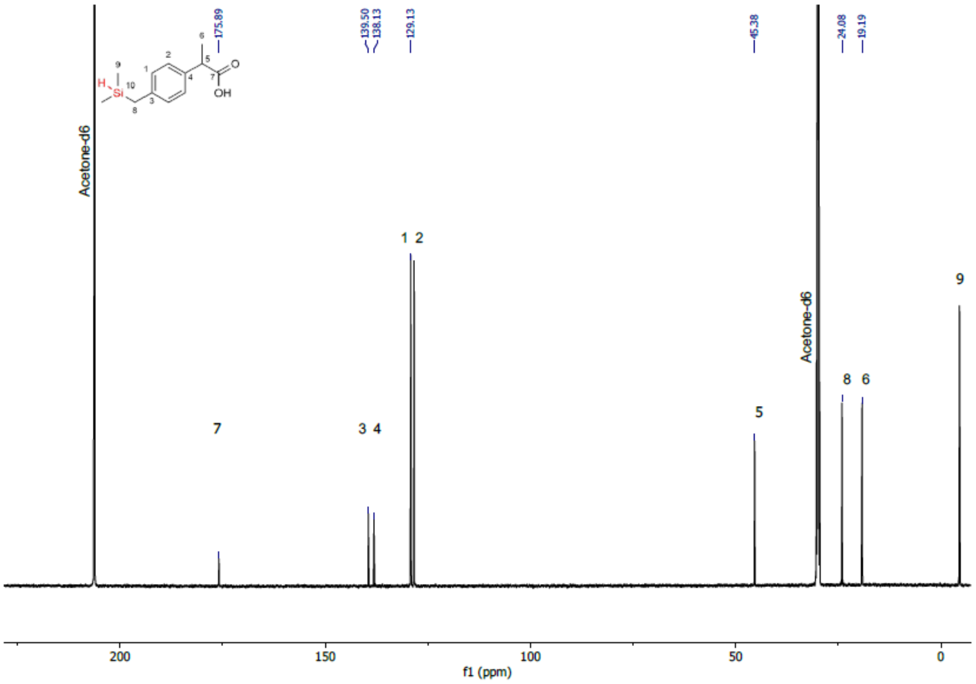


Figure 5.4: $^{13}C\{^1H\}$-NMR of sila-ibuprofen measured in acetone-d6 at 151 MHz. Signals are assigned numbers according to labeling scheme in top left corner.

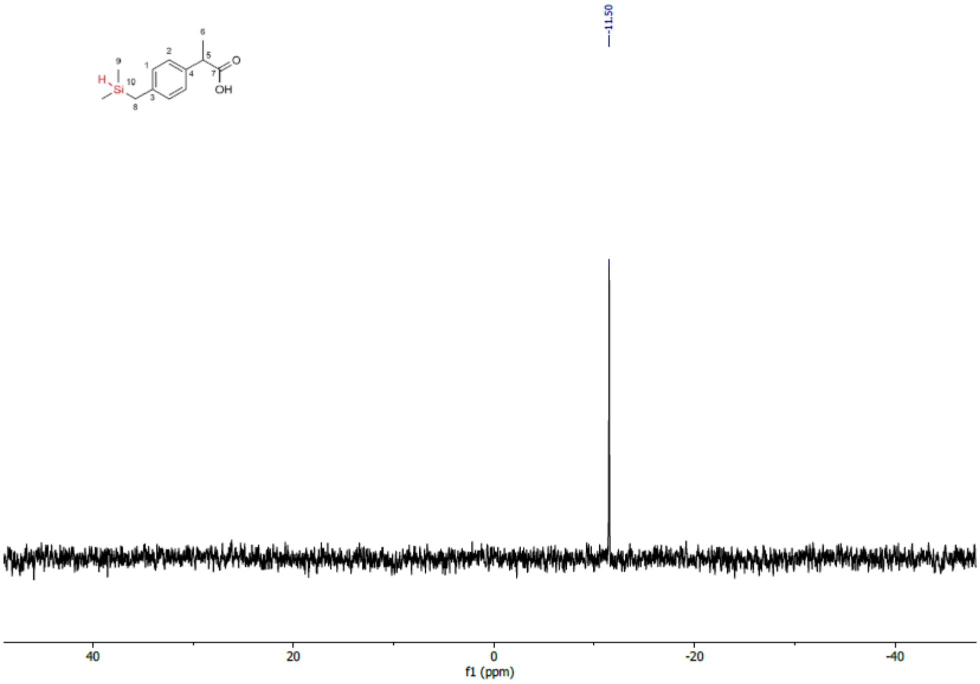


Figure 5.5: $^{29}Si\{^1H\}$-NMR of sila-ibuprofen measured in acetone-d6 at 119 MHz.

The oxidation product of the reaction of sila-ibuprofen with a Pearlman's catalyst is the silanol, where the hydrogen atom at the silane functional group of sila-ibuprofen is exchanged with a hydroxy group. The structure and corresponding NMR spectra are shown in Figure 5.6 to 5.8.

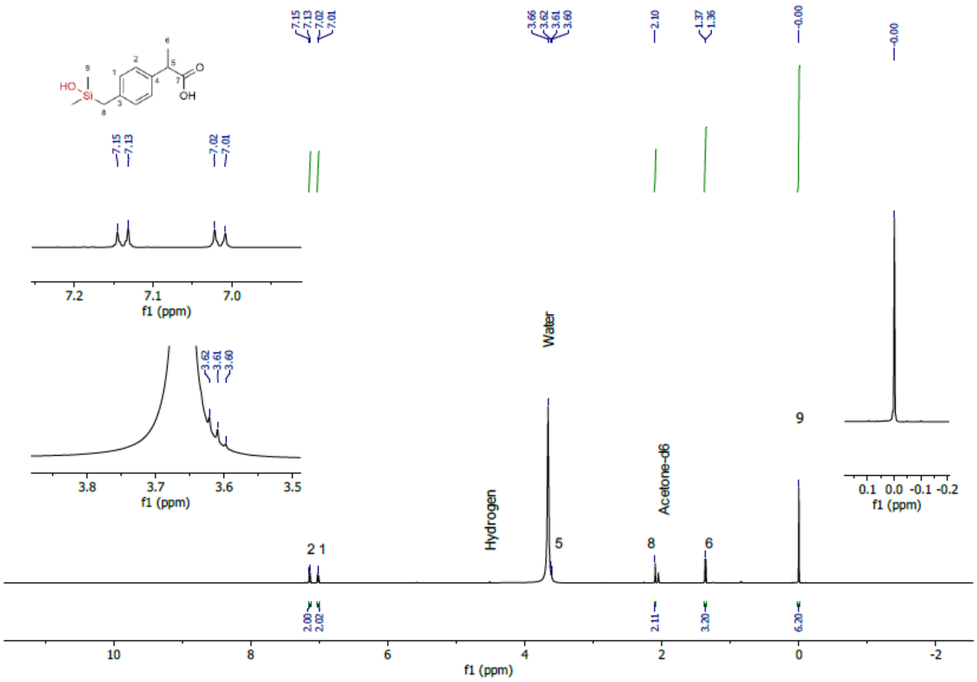


Figure 5.6: ^{1}H-NMR of hydroxy-sila-ibuprofen measured in acetone-d6 at 600MHz. Signals are assigned numbers according to labeling scheme in top left corner.

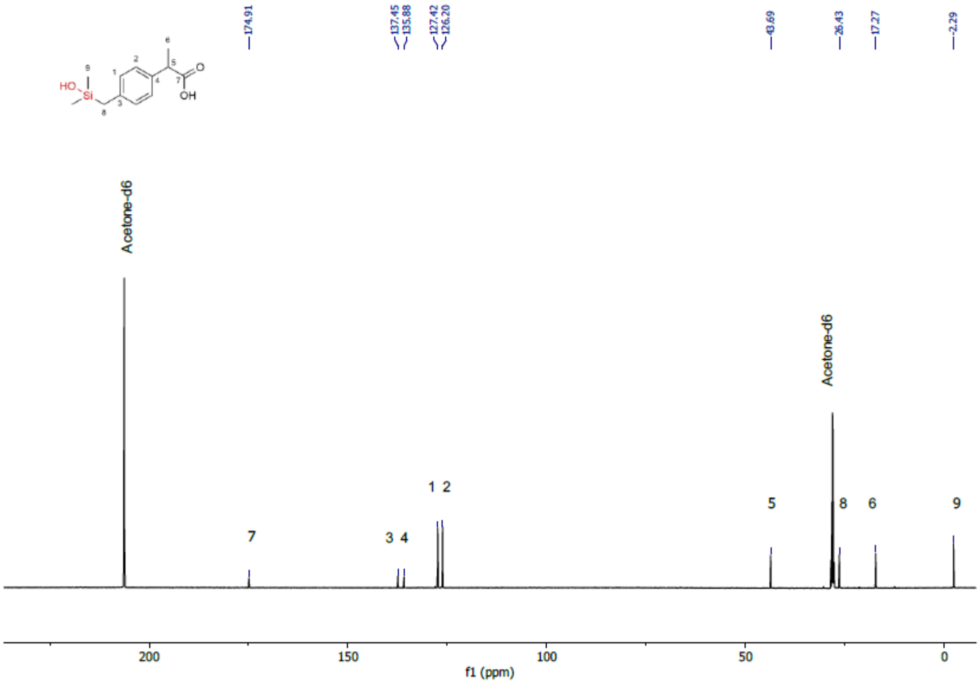


Figure 5.7: ^{13}C{^{1}H}-NMR of hydroxy-sila-ibuprofen measured in acetone-d6 at 151 MHz. Signals are assigned numbers according to labeling scheme in top left corner.

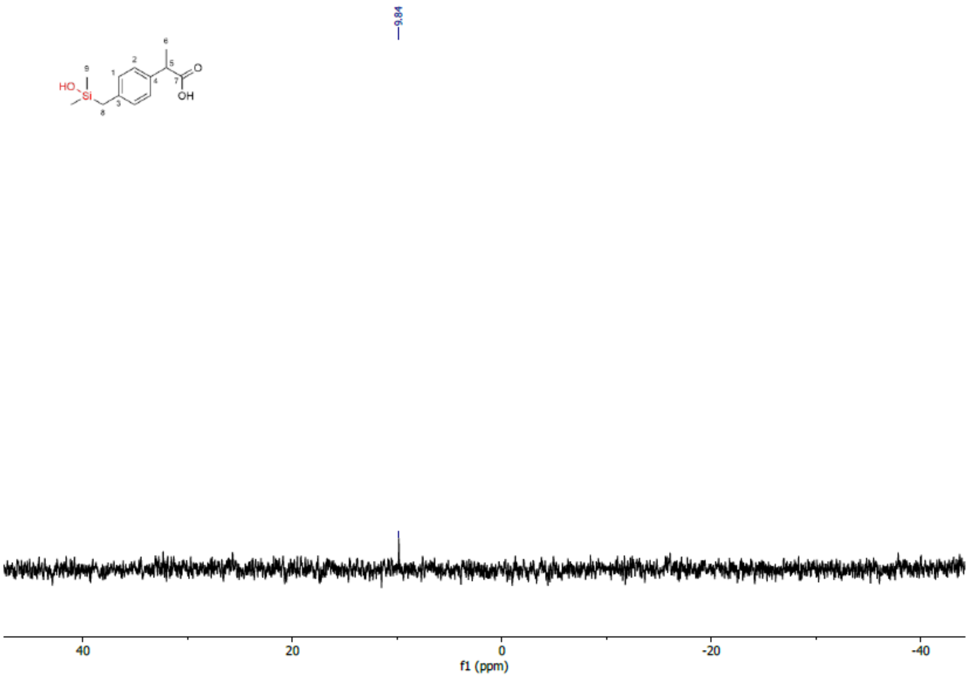


Figure 5.8: ^{29}Si{^{1}H}-NMR of hydroxy-sila-ibuprofen measured in acetone-d6 at 119 MHz.

In addition to the oxidation product, it was observed that upon removal of the solvent the silanol forms the dimer siloxane through elimination of water from two hydroxyl functions. This dimer was impossible to isolate, since it also formed the silanol again in an equilibrium, but NMR-spectra of this mixture were measured and are shown in Figures 5.9 and 5.10.

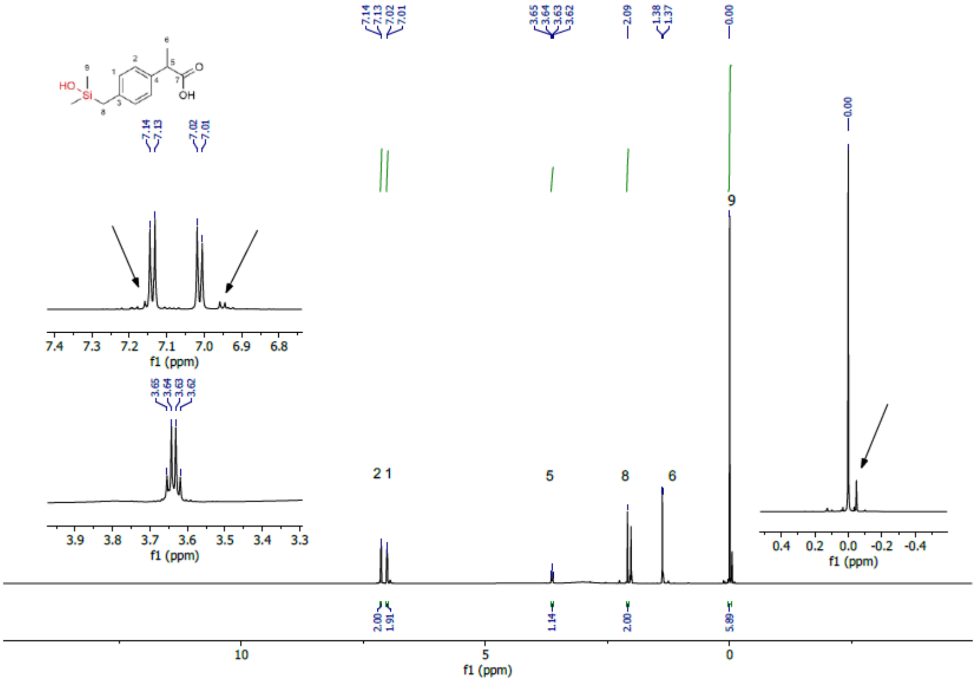


Figure 5.9: ^{1}H-NMR of hydroxy-sila-ibuprofen with removed solvent forming the siloxane in equilibrium measured in acetone-d6 at 600MHz. Signals are assigned numbers according to labeling scheme in top left corner. Signals of siloxane highlighted with arrows.

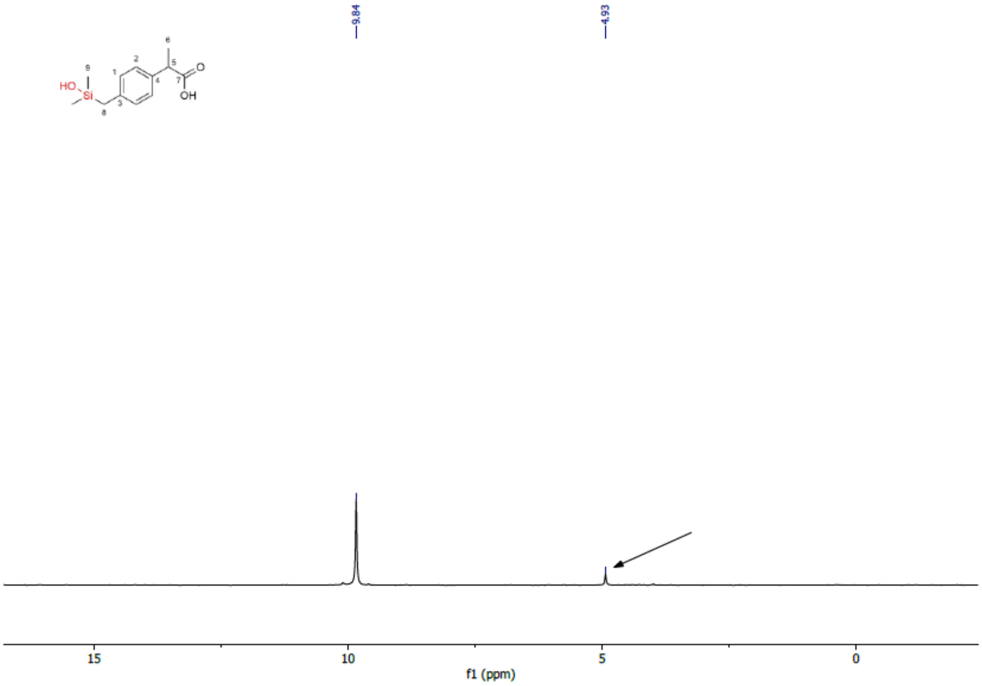


Figure 5.10: ^{29}Si{^{1}H}-NMR of hydroxy-sila-ibuprofen with removed solvent forming the siloxane in equilibrium measured in acetone-d6 at 119 MHz. Signal of the siloxane marked with arrow.

5.2.2.4 Stability in Aqueous Solution

Since silanes as a class of substances tend to be very oxophilic and unstable at ambient conditions, the question of stability of sila-ibuprofen in aqueous media needed to be addressed, especially if a use in biological media and patients is the final goal of the development. There are two possible pathways for sila-ibuprofen to decompose in solution. Either it forms the silanol by oxidation of the Si-H bond which was characterized by NMR in the previous chapter 5.2.2.3 or it cleaves the Si-C bond which would leas to significantly smaller metabolites. To perform tests on the time scale the decomposition of sila-ibuprofen shows, a NMR-tube of sila-ibuprofen dissolved in D_2O with 0.9 % of NaCl was measured every 7 days. An identical solution, that was stored at 4°C between measurements, was measured every 28 days. These spectra were measured with the help of Daniel Duvinage. The resulting spectra are shown in Figures 5.11.

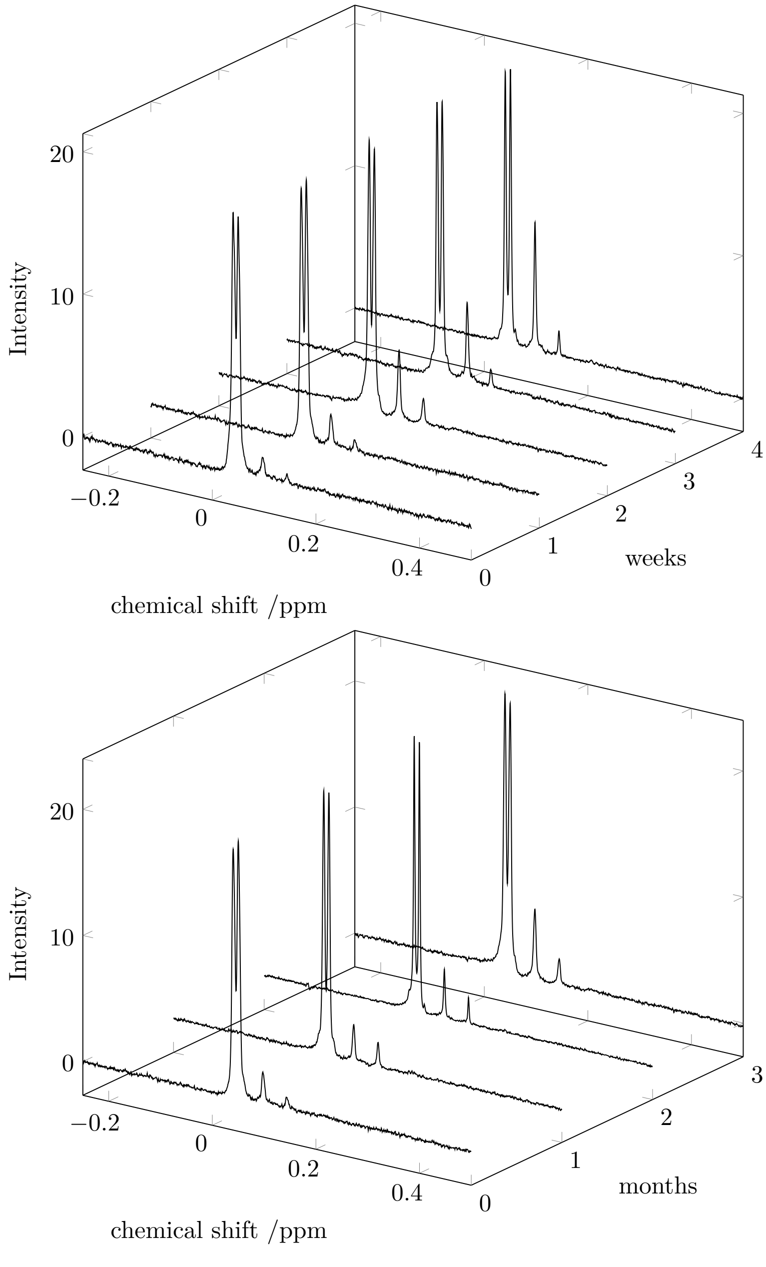


Figure 5.11: Stability of sila-ibuprofen monitored by NMR spectra measured in D_2O with 0.9 % NaCl in intervals of 7 days stored at room temperature (top) and in intervals of 28 days stored at 4°C (bottom).

The plots in Figure 5.11 show the peaks of the methyl protons next to the silicon atom and the surrounding spectral region. It is clearly visible, that at room temperature after 4 weeks a slow decomposition began and side products show a signal at ca. 0.07 ppm. Additional smaller amounts of a new substance are observed at ca. 0.11 ppm. These two decomposition compounds could be the silanole as the first bigger signal and then the dimeric siloxane, since it shows a similar behavior of significantly smaller ratio compared to the siloxane as observed in the spectra in Figures 5.9 and 5.10. The same trend can be observed during storage at 4 °C over months, which shows that a fresh solution of the material should be stable enough over a period of time, where it could be used in practical applications. However, long term storage of prepared solutions might be difficult.

Additional investigations were performed where sila-ibuprofen was stored in solutions under basic conditions adding $NaHCO_3$ at pH 8, 10 and 12. At physiological pH (pH = 8) no decomposition was observable over the course of one month. At pH 10 and 12 sila-ibuprofen formed the sodium salt of hydroxy-ibuprofen under the release of hydrogen gas. This ensures that even in the basic conditions of the digestive system the drug is not decomposing before it is resorbed.

5.2.3 Crystallography of Sila-Ibuprofen

The chemical characterization was readily possible after synthesis except for X-ray diffraction experiments, as it proved very difficult to obtain sufficiently crystalline material by conventional crystallization techniques. Most attempts to crystallize the material were performed using combinations of solvents or evaporation of them. Since they usually result in oily mixtures, the only successful method to obtain crystals was the resublimation of the substance, however over a time-span of many months, which was inconvenient for repetition of experiments. The use of high vacuum with slightly elevated temperature in a water bath and the use of a resublimation apparatus was finally able to produce crystalline material suitable for diffraction experiments.

Diffraction experiments were performed at the synchrotron facility SPring-8 in Hyogo, Japan at beamline BL02-B1 using a curved image plate detector. This radiation source was chosen to obtain the highest possible resolution and data quality for the development of force field parameters and comparison of properties. The results after refinement of the measured data are given in Table 5.1.

Ibuprofen and sila-ibuprofen form isostructural crystals in the space group $P\ 2_1/c$ with very similar lattice parameters. The refinement statistics are the results after XWR, that is initial solution using *shelxT*, IAM refinement using *olex2.refine* and subsequent XWR, consisting of HAR on a level of theory of B3LYP/def2-TZVPP using a radius of 8 Å for the embedding in cluster charges and a dimer of the molecules formed by the hydrogen bond between the carboxylic acid group grown as the wavefunction part using the *Olex2*-GUI in the HARt implementation. The XCW-step was performed on a level of theory of HF/def2-TZVPP without any cluster charges in order to retrieve all experimentally available information without the bias of an overestimation through a selection of the method (compare section 4.1). The final structures are shown in Figure 5.12.

Table 5.1: Refinement statistics of ibuprofen and sila-ibuprofen after XWR using *Tonto*. [153]

Structure	Ibuprofen	Sila-ibuprofen
Space group	P 2_1/c	
a /Å	14.465(3)	14.814(3)
b /Å	7.815(2)	7.972(2)
c /Å	10.435(2)	10.798(2)
β /°	99.66(3)	100.70(3)
V /Å^3	1162.9(4)	1253.0(4)
T /K	25	
d_{min} /Å	0.45	
$\lambda_{\text{X-ray}}$ /Å	0.3567	
R_{int}	0.0552	0.0557
Avg. redundancy	9.49	6.71
Completeness	1.00	1.00
Average I/σ	41.3	22.3
# of refln. measured	132378	105581
# of unique refln.	13367	14320
Obs. Criterium	$F_o^2 > 2\sigma(F_o^2)$	
# of observed refln.	10549	10780
Weighting scheme	$w = 1/\sigma(F_o)$	
λ_{XWR}	0.40	0.63
χ^2	1.7817	2.5273
Final R_1	0.0216	0.0267
Final wR_2	0.0285	0.0365
Max $\Delta\rho$ /eÅ^{-3}	0.183	0.261
Min $\Delta\rho$ /eÅ^{-3}	-0.139	-0.192
CCDC deposition number	1983628	1983627

5.2.4 Complementary Bonding Analysis

Based on the wavefunctions obtained from XWR, a comparison of the bonding situations in both molecules was performed employing the methods mentioned in section 1.3, namely the frameworks of QTAIM, NBO and ELI-D in this case. Results are shown for selected bonds in Figure 5.12.

The bonds in the carboxylic acid group, corresponding to the anchor inside the enzyme where ibuprofen exhibits its main function, are mostly unchanged. In both ibuprofen and sila-ibuprofen the QTAIM-, NBO- and ELI-based bond descriptors as well as the atomic charges reflect very similar behavior and show that both acids should have similar acidity and electrostatic properties based on the assumption that the charge distribution mainly dictates the electrostatic properties. In the bonds around the carbon/silicon switched position, the effect of the *umpolung* (see section 1.7) is highly visible. The charge of the hydrogen atom attached to the ternary position is switched from around +0.2 e in ibuprofen (NPA charges) to -0.2 e in sila-ibuprofen, and from approx. 0 e to -0.7 e in the QTAIM description. Also, the charge of the carbon/silicon is switched in sign: while the carbon analogue shows 0.1 e and -0.2 e in QTAIM/NPA, the silicon atom has 2.8 e/1.4 e charge on this silicon atom.

Moreover, the nature of the bonds between C-H/Si-H changes: while in the QTAIM picture the density, its Laplacian and the delocalization index (DI) hint towards a highly covalent single bond, the situation is quite different in the silane function, where the Laplacian is positive and the DI is below 0.5. The Raub-Jansen index (RJI) decreased significantly, too, pointing towards a much higher shift of the bond polarization towards the hydrogen atom. The same trend is observed for the C-C/Si-C bonds, shown as an example for the methylene group. The C-C bond is much more covalent in ibuprofen (high density, negative Laplacian, DI around 1, 95% covalency in NRT and 50% RJI) while the Si-C bond is more polarized (lower density, positive Laplacian, DI

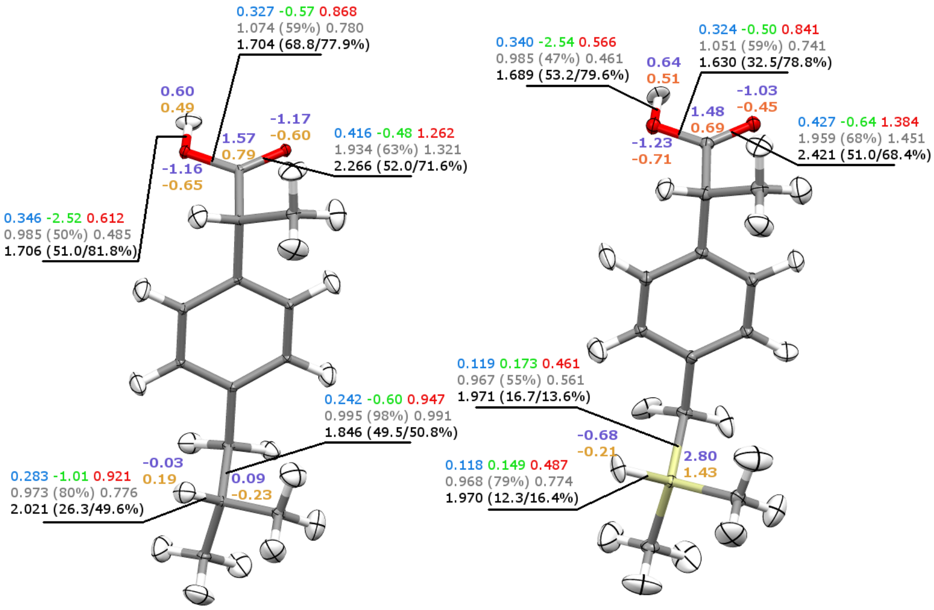


Figure 5.12: Final crystal structure of ibuprofen (left) and sila-ibuprofen (right) after XWR showing ADPs at 80% probability level with labeled bond descriptors, using the following color code: QTAIM charges in e, NPA Charges in e, ρ and $\triangle\rho$ at the BCP in a.u., delocalization index, NBO bond order with percentage weight of covalent form in parenthesis, NLMO/NPA Bond order and population of the diatomic ELI-D basin with contribution of electron density from the element with higher Z into the bond by volume/electron density, respectively (black).

around 0.5, 55% covalency in NRT reflected by low covalent NBO-BO and quite polarized RJI below 17% in both cases).

These observations allow the assumption that the biologically active site, the carboxylic acid function was unaffected by the carbon-silicon exchange while some new bonding situation and especially charge distribution was achieved in the formerly unpolar tert-butyl group. This will further be discussed in section 5.3.2.4.

5.2.5 Electrostatic Potential

To understand the effect of the changed charges on the exhibited electrostatic potential between ibuprofen and sila-ibuprofen is was plotted onto the Hirshfeld suface of the molecules within the crystal structure shown in Figure 5.13. It is observable that around the silane function in sila-ibuprofen exhibits a strong negative potential while ibuprofen has a weak positive potential at the corresponding position. This is explained by the relatively high negative charge of the hydrogen atom in sila-ibuprofen in comparison to the relatively neutral hydrogen atom in ibuprofen.

The wavefunction obtained from the XWR refinements were also used for further analysis of the ESP on the Hirshfeld surfaces by tricking *CrystalExplorer* to read them from prepared *.fchk* files after conversion from *Tonto* to *.fchk* files using *cuQCrT* (compare section 3.3). To understand the difference of packing and electrostatics in these two in principle isomorphous compounds, the Politzer parameters were calculated for these surfaces and resulting values are summarized in Table 5.2. [467]

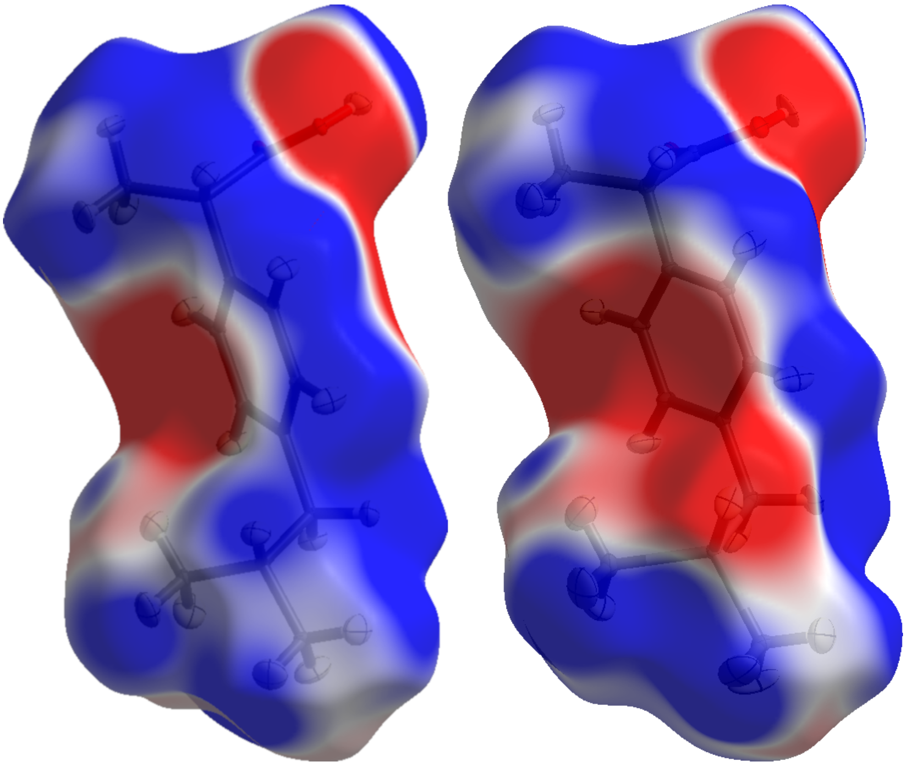


Figure 5.13: Electrostatic Potential (ESP) mapped onto the Hirshfeld surface of ibuprofen (left) and sila-ibuprofen (right) from the XCW wavefunctions. Color scale from -0.01 a.u. (red) to 0.01 a.u. (blue). Visualized using *CrystalExplorer*. [423]

Table 5.2: Politzer parameters [467] of ibuprofen and sila-ibuprofen after XWR using *Tonto*, importing the wavefunction into *CrystalExplorer* using a script which copies the *.fchk* files to the according places. [153,423] Units in $e\text{Å}^{-1}$ and $e^2\text{Å}^{-2}$. V$_+$: Mean positive values, V$_-$: Mean negative Values, Π: average deviation from mean value on the surface, σ_+^2, σ_-^2 and σ_{Tot}^2: variance of surface values (positive, negative and all values) and ν electrostatic balance (max=0.25).

	Ibuprofen	Sila-ibuprofen
V$_+$	0.0217	0.0246
V$_-$	-0.0306	-0.0302
Π	0.0240	0.0263
σ_+^2	63548	78665
σ_-^2	36953	52630
σ_{Tot}^2	100501	131295
ν	0.2325	0.2401

The parameters can be interpreted in order to see differences in deviations from the zero-potential and the spatial distribution on the surface. V$_+$ and V$_-$ give rise to the average value of the respective potential regions. If there is a higher average value of the potential in either direction, the molecular potential can be interpreted as more pronounced. Π, being the average deviation from the mean value of the surface, gives rise of the spread of values. If it is low all values are more similar than if Π is high. σ^2_+ and σ^2_- are the variance of the values on the surface where the value has respective sign, σ^2_{Tot} is the variance of all values. ν refers to the balance of the potential distribution, where 0.25 is the maximal value by definition. If the value is lower the potential is not balanced over the complete surface. In the comparison of ibuprofen and sila-ibuprofen, the potential expressed by ibuprofen is less balanced, as $\nu = 0.2325$, which coincides with V$_+$ and V$_-$ being further apart than in sila-ibuprofen, where $\nu = 0.2401$.

Quantification of the effect of the *umpolung* in the silane function is available when comparing the dipole moment of model compounds trimethylsilane, as used in section 5.2.6, and the carbon equivalent tert-butane.

Both molecules were geometry optimized on a level of theory of B3LYP/def2-TZVPP and the dipole moment along the C-H/Si-H bond calculated. The magnitude of the dipole moment of tert-butane is 0.0528 a.u. along the C-H bond, while in trimethylsilane the dipole moment is -0.2171 a.u. along the Si-H bond. The opposing sign is kept to denote the opposite orientation of the dipole moment. This shows how this carbon-silicon switch introduces an almost 4-times higher dipole moment in the direct proximity while changing the direction of the dipole moment, as well.

5.2.6 Force Field Development

The force field parameters for the MD simulation of ibuprofen inside their targets COX-I and COX-II were taken from default CHARMM force fields and the online service *swissparam*. [235–237, 468] Here, coordinates of a small molecule can be uploaded and parameters of a databank of force fields are given for use in a simulation. Since the carbon-silicon bond found in sila-ibuprofen is a rare type of bond in the context of protein molecular dynamics simulations, there was a lack of parameters for the description of sila-ibuprofen. The missing bonded parameters are visualized in Figure 5.14. In addition, the non-bonded parameters of the atom types Si, H_{Si} and the neighbouring carbon atoms needed to be modelled.

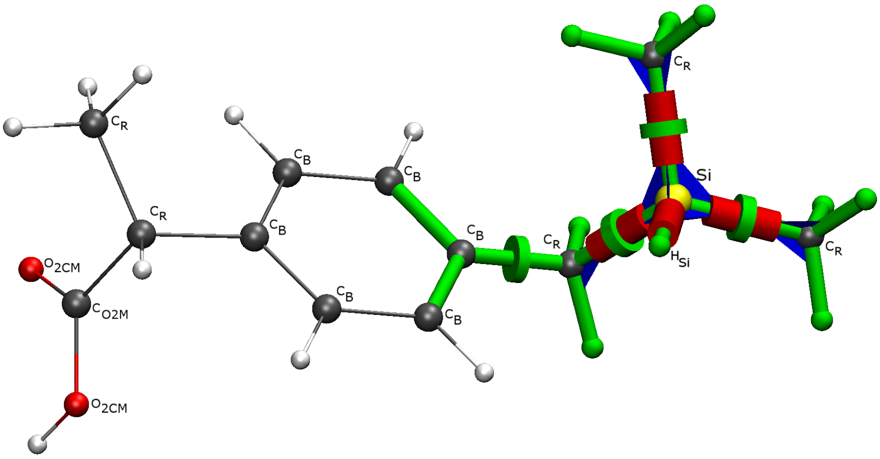


Figure 5.14: Visualization of missing force field parameters for molecular dynamics simulations of sila-ibuprofen. Red cylinders denote missing bond-, blue triangles missing angle- and green discs missing dihedral-parameter. All hydrogen atoms without labels are of type H_{CMM}. Visualized using VMD and ffTK. [56, 469]

A convenient tool inside the GUI *VMD*, [56] called *ffTK*, the force field Tool Kit, guides the user through the process of optimizing a force field of a small molecule based on a series of calculations.

The calculation of the forces between atoms in the XWR derived wavefunction was carried out, reading it into *Gaussian09* similarly to a NBO calculation, then calculating numerical gradients and Hessian matrices. Using these for the estimation of force constants for bonded force parameters between two atoms and angle force parameters between three atoms the missing parameters were estimated. Dihedral barriers were optimized using a potential energy surface scan on a level of theory of MP2/def2-SVPP rotating around the axes under investigation (compare green discs in Figure 5.14). [344] A plot of the energies of dihedral parameters and the quantum mechanical energies for the corresponding conformations during the PES-scan are shown in Figure 5.15.

The last missing parameters were charges of atoms. They were estimated using the possibility of *fftTK* to set up calculations of dimers of water with the molecule under investigation in order to match the Coulomb energy of the charges in the force field with those of the QM calculation. A level of theory of HF/def2-TZVP was employed. [344] This procedure was performed for all atom types under the constraint that the total charge of the deprotonated sila-ibuprofen molecule is expected to have an overall charge of -1 *e*.

The final challenge was the optimization of Lennard-Jones parameters to correctly describe the dispersion interactions. To tackle this problem, a dimer of a model compound (trimethylsilane) was used, once arranged as

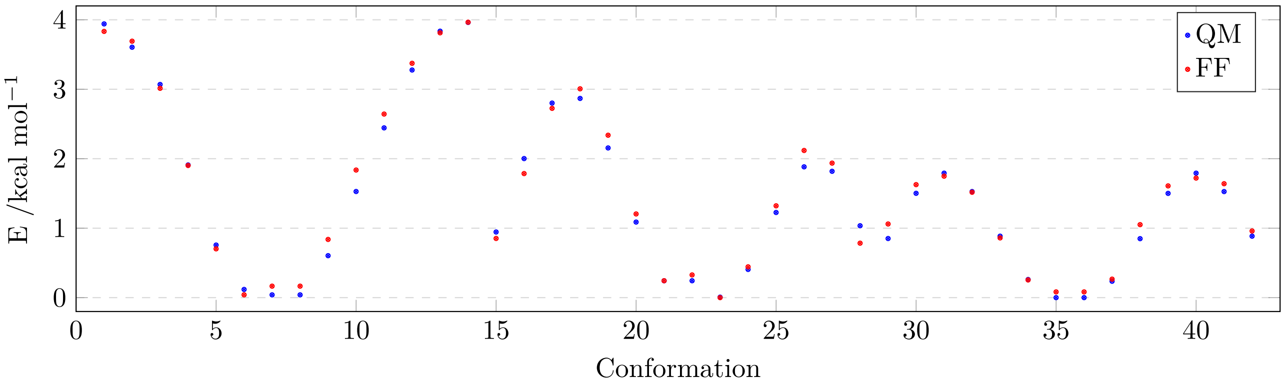
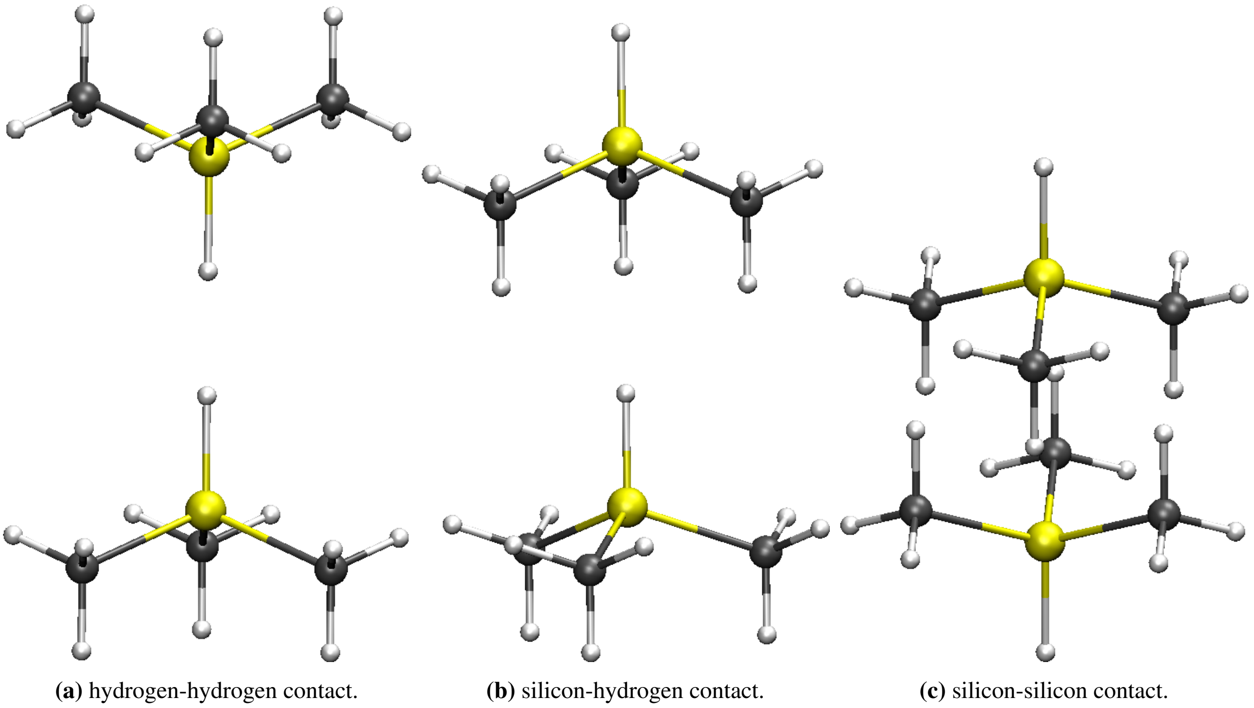


Figure 5.15: Plot of the energy derived by the PES scan using Quantum Mechanics (QM) and the optimized Force Field (FF) parameters. Each point on the principal axis corresponds to a different geometry.

a hydrogen-hydrogen contact, once as silicon-silicon and once as hydrogen-silicon contact dimer. The positions of molecules are visualized in Figure 5.16.



(a) hydrogen-hydrogen contact.

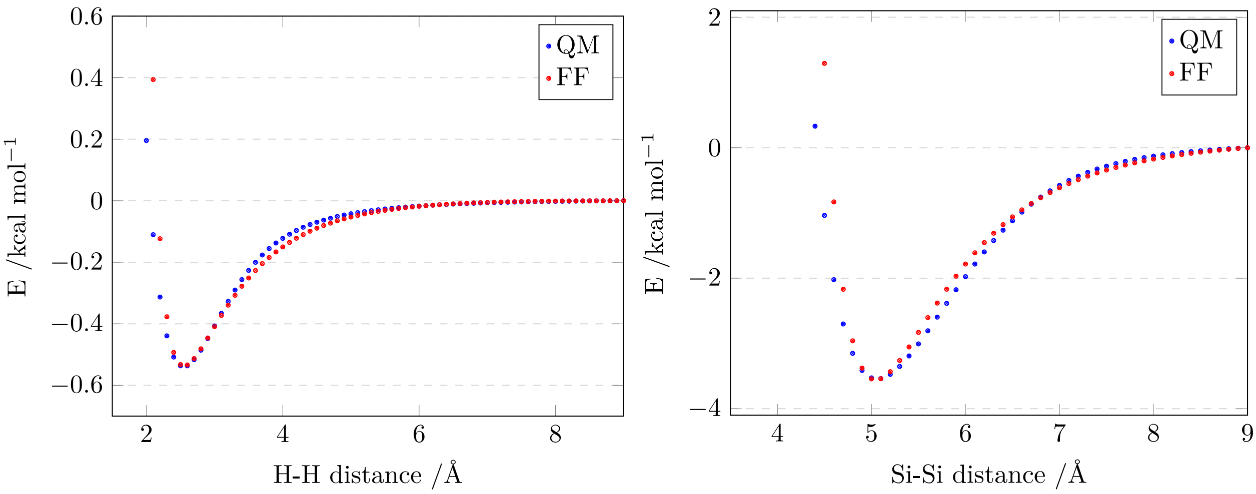
(b) silicon-hydrogen contact.

(c) silicon-silicon contact.

Figure 5.16: A visualization of the arrangement of the trimethylsilane dimers for development of the Lennard-Jones parameters of the force field. Visualized using VMD. [56]

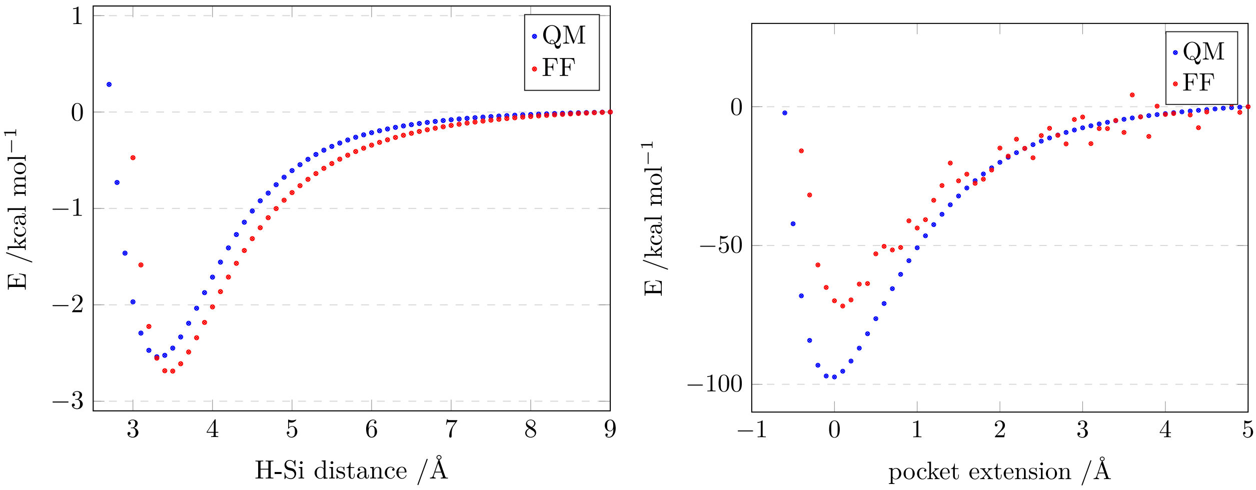
These arrangements were used for restricted PES-scans using *ORCA* for the quantum mechanical calculation of energies on a level of theory of B3LYP-D3BJ/def2-TZVP. [344, 399, 400] It is extremely important to use the dispersion correction in the quantum mechanical calculation to be able to fit the parameters of the force field reasonably. The geometries obtained from the wavefunction calculation, when shortening the corresponding contact in 0.1 Å, steps was pasted into a xyz-file to be used as the input for the calculation of energies using *NAMD2*. The energies were plotted and the agreement calculated in terms of least squares between the point of the quantum mechanical energy and the force field energy. The energy agreement is plotted in Figures 5.17a to 5.17c.

To ensure physical behaviours of the force field in the enzyme pocket, the geometry of the enzyme pocket from the COX-II entry 4PH9 in the Protein Database [470] was used to set up a calculation of all amino acids with atoms in a distance of less or equal to 5 Å to any atom of ibuprofen in the active site of subunit 1 of COX-II. The center of mass of each amino acid was used to push them away in an anti-concentric motion from the center of mass of ibuprofen in a PES-Scan. The level of theory used was B3LYP-D3BJ/def2-SVPP, [344, 399, 400] accommodating a big number of atoms in the quantum mechanical part. Identical geometry was used to calculate the energy based on the derived force field to see whether the position and trend of the pocket size and position can be reconstructed. To compensate the difference of geometry when using sila-ibuprofen in the pocket, the bonds of changing atoms and their neighbours were reset to the equilibrium distance of the derived bonded parameters of the force field while retaining angles and dihedral angles using internal coordinates. The results of this scan are shown in Figure 5.17d. Unfortunately, the output of the energy obtained by *NAMD2* has limited precision, therefore the plot scatters a lot.



(a) hydrogen-hydrogen contact.

(b) silicon-silicon contact.



(c) hydrogen-silicon contact.

(d) extension of enzyme pocket.

Figure 5.17: Plot of quantum mechanically derived and force field computed energies for trimethlysilane and sila-ibuprofen in the active site of COX-II.

The plots show that within reasonable accuracy the positions and depths of minima of all models are reproducible using the force field. The simulation of the pocket extension shows a good agreement of the minimum of the energy when altering the distance of amino acids. The depth is not exactly reproduced, but it is difficult to say whether this is due to the inaccuracy of the Lennard-Jones parameters of the new silicon and hydrogen atom types, or rather the inaccuracy of the other types, e.g. the amino acids as well as the remaining atoms of the *swiss param* service. Therefore, these parameters were accepted and the parameters summarized in Tables 5.3 to

5.5 were used for the following simulations.

Table 5.3: Force field parameters for bonds (top) and angles (bottom) used for ibuprofen and sila-ibuprofen.

Bond	Force constant (K_b) /kcal mol^{-1} Å^{-2}	Eq. distance (b_0) /Å
C_R Si	190.6411	1.8677
Si H_{Si}	171.6976	1.4972
C_{O_2M} O_{2CM} ^{a}	702.1030	1.2610
C_R C_R ^{a}	306.4320	1.5080
C_R H_{CMM} ^{a}	342.9910	1.0930
C_R C_B ^{a}	356.7370	1.4860
C_R C_{O_2CM} ^{a}	275.6310	1.5100
C_B C_B ^{a}	401.0680	1.3740
C_B H_{CMM} ^{a}	381.8530	1.0840

Angle	Force constant (K_θ) /kcal mol^{-1} rad^{-2}	Eq. angle (θ_0) /rad
C_B C_R Si	30.2479	105.4005
C_R Si H_{Si}	22.5564	116.2359
C_R Si C_R	54.1973	114.2112
Si C_R H_{CMM}	14.9964	110.2184
C_B C_B C_B ^{a}	48.1450	119.9770
C_B C_B C_R ^{a}	57.7880	120.4190
C_B C_B H_{CMM} ^{a}	40.5170	120.5710
C_R C_R H_{CMM} ^{a}	45.7700	110.5490
H_{CMM} C_R H_{CMM} ^{a}	37.1340	108.8360
C_B C_R C_R ^{a}	54.4060	108.6170
C_B C_R H_{CMM} ^{a}	45.1220	109.4910
C_B C_R C_{O_2CM} ^{a}	71.9660	109.5000
C_R C_R C_{O_2CM} ^{a}	23.7490	98.4220
H_{CMM} C_R C_{O_2CM} ^{a}	37.7820	108.9040
C_R C_R C_R ^{a}	61.2430	109.6080
C_R C_{O_2CM} O_{2CM} ^{a}	87.0070	114.6890
O_{2CM} C_{O_2CM} O_{2CM} ^{a}	84.9910	130.6000

[a] Parameters taken from *swissparam*. [468]

Table 5.4: Force field parameters for dihedral angles and impropers used for ibuprofen and sila-ibuprofen.

Dihedral	Force constant (K_χ) / kcal mol^{-1} (K_ψ) / kcal mol^{-1} rad^{-2}	Multiplicity (n)	Eq. dihedral (δ) / °
C_B C_B C_R Si	0.442	2	180.00
C_B C_R Si C_R	0.331	3	0.00
C_B C_R Si H_{Si}	1.091	1	0.00
C_R Si C_R H_{CMM}	0.000	3	0.00
H_{CMM} C_R Si H_{Si}	0.301	3	0.00
C_B C_B C_B C_B [a]	3.500	2	180.00
C_B C_B C_B H_{CMM} [a]	3.500	2	180.00
C_B C_R C_R H_{CMM} [a]	0.195	3	0.00
C_B C_R C_{O_2CM} O_{2CM} [a]	0.300	2	180.00
C_B C_B C_R C_R [a]	0.225	2	180.00
C_B C_B C_R H_{CMM} [a]	-0.210	2	180.00
C_B C_B C_R H_{CMM} [a]	0.196	3	0.00
C_B C_B C_R C_{O_2CM} [a]	0.100	3	0.00
C_B C_B C_B C_R [a]	3.500	2	180.00
C_BC_R C_R C_R [a]	0.150	3	0.00
C_R C_R C_{O_2CM} O_{2CM} [a]	0.631	2	180.00
C_R C_B C_B H_{CMM} [a]	3.500	2	180.00
H_{CMM} C_R C_R H_{CMM} [a]	0.142	1	0.00
H_{CMM} C_R C_R H_{CMM} [a]	-0.693	2	180.00
H_{CMM} C_R C_R H_{CMM} [a]	0.157	3	0.00
H_{CMM} C_R C_{O_2CM} O_{2CM} [a]	-0.053	3	0.00
C_R C_R C_R H_{CMM} [a]	0.320	1	0.00
C_R C_R C_R H_{CMM} [a]	-0.315	2	180.00
C_R C_R C_R H_{CMM} [a]	0.132	3	0.00
H_{CMM} C_B C_B H_{CMM} [a]	3.500	2	180.00
C_{O_2CM} C_R C_R H_{CMM} [a]	-0.070	3	0.00
C_B C_B C_B C_R [a]	2.879	-	-
C_B C_B C_B H_{CMM} [a]	1.079	-	-
C_B C_R C_B C_B [a]	2.879	-	-
C_{O_2CM} O_{2CM} C_R O_{2CM} [a]	12.810	-	-

[a] Parameters taken from *swissparam*. [468]

Table 5.5: Force field parameters for Lennard Jones potentials (top) and charges of atoms in *e* (bottom) used for ibuprofen and sila-ibuprofen.

Atom type	ϵ /kcal mol^{-1}	R$_{min}$ /Å
C$_B$ ^{a}	-0.0700	1.9924
C$_R$ ^{a}	-0.0550	2.1750
H$_{CMM}$ ^{a}	-0.0220	1.3200
C$_{O_2M}$ ^{a}	-0.0700	2.0000
O$_{2CM}$ ^{a}	-0.1200	1.7000
H$_{Si}$	-0.0152	1.5210
Si	-0.5650	2.3800

Atom type	Ibuprofen^a	Sila-ibuprofen
O$_{2CM}$	-0.9000	-0.9000
C$_B, q$	-0.1435	-0.1435
C$_{B,t}$	-0.1500	-0.1500
C$_{O_2M}$	0.9060	0.9060
C$_{R,CH_2}$	0.1435	0.1480
C$_{R,CH}$	0.0375	0.0375
C$_{R,CH_3-C/Si}$	0.0000	0.0000
H$_{C_R}$	0.0000	0.0000
H$_{C_B}$	0.1500	0.1500
H$_{CH/HSi}$	0.0000	-0.1470
CH/Si	0.0000	0.1430

^{a} Parameters taken from *swissparam*. [468]

5.2.7 MD Settings

MD simulations were performed using *NAMD2* [471, 472] and a CHARMM36 force field, where missing parameters for small molecules were obtained from *swissparam*, if available. [235–237, 468]
Preparation of input files: The structures of the ibuprofen-COX-I complex [473] and ibuprofen-COX-II complex [470] as deposited in the PDB were downloaded, a solvation box of 15 Å of water molecules and a physiological concentration of NaCl added around the protein using VMD plugins. [56]
Settings for the production runs: The target temperature of the barostat was set to 300 K, the pressure to 1.01325 bar. The thermostat used was a Langevin coupling, which was decoupled from the hydrogen atoms and had a damping factor of 1. The barostat was an isotropic Langevin-Piston mechanism, which had a period of 200 steps, decay time of 100 steps and a target temperature of 300 K. The integration frequency was 2 *fs*, the full electrostatics were also re-evaluated every 2 fs. The non-bonded interactions that are not in the full Ewald summation, were evaluated every *fs*. The cut-off for non-bonded interactions was 12 Å where at 10 Å a switching function was used to smoothen the transition to 0. The geometry, velocities and electrostatics were saved every *ps*, the pressure and energies printed every 80 *fs*. Pre-compiled binaries of *NAMD2* with *CUDA* acceleration were used for simulations, except for FEP, where *CUDA* acceleration is not yet implemented. The normal strategy was a minimization of the structure obtained from the preparation procedure then slowly annealing the temperature from 60 K in 1 K steps with 600 simulation steps of 2 *fs* each until 300 K were reached. A subsequent equilibration simulation with production settings and 500,000 steps was appended to ensure equilibrium of the complex. Then the production simulation was started. The starting geometry for the runs of sila-ibuprofen were generated by using the deposited ibuprofen complexes and manually changing the ibuprofen residues into sila-ibuprofen.

5.2.8 MD Simulations

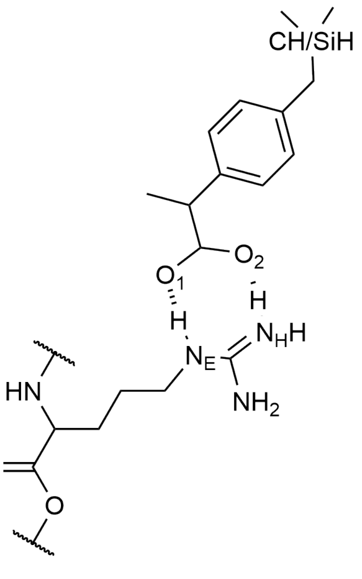


Figure 5.18: Lewis structure for the ibuprofen/sila-ibuprofen - arginine contact as the anchoring motif in COX-I and COX-II with name scheme of atoms.

A force field for the study of similarities or differences in bonding of ibuprofen and sila-ibuprofen inside the active side of COX-I and COX-II was derived as described in section 5.2.6 and used for simulations over a timespan of 40 ns. The geometry was analyzed over this timespan to understand the similarity or difference in bonding of the molecules and whether sila-ibuprofen would prove to be similarly active as a reversible inhibitor as ibuprofen. The reported key bonding motif of ibuprofen in the active site of COX enzymes is a two-fold hydrogen bond between the carboxylic acid group of ibuprofen and an arginine residue, [299] as depicted in Figure 5.18. A ring of 8 atoms, where the O-H-N contact is almost linear, is formed. This contact is the same as used for arachidonic acid during the normal function of COX and therefore inhibits the active site completely. This bonding motif is stated to be the leading contribution for the binding of ibuprofen in the active site. Therefore, the stability of this contact was monitored by plotting the distance between oxygen and nitrogen atoms of the respective residues over the production runs. The resulting plots are shown in Figures 5.19 and 5.20, the first being based on simulations for COX-I and the latter being the analysis of simulations inside COX-II. There are two contacts of each kind, since COX-enzymes consist of 2 subunits. This means there are two strands of protein that are connected to form one entity with two active sites. This does not need to be the case for proteins consisting of multiple subunits, like for example in ribosomes where the two subunits (small and big subunit) only form the active entity upon combination. In the case of COX the two subunits have identical gene sequence and the two isoenzymes COX-I and COX-II have 65% of the genetic sequence in common. [474–477]
The numbers for the subunits were assigned based on the appearance on the PDB files downloaded from the

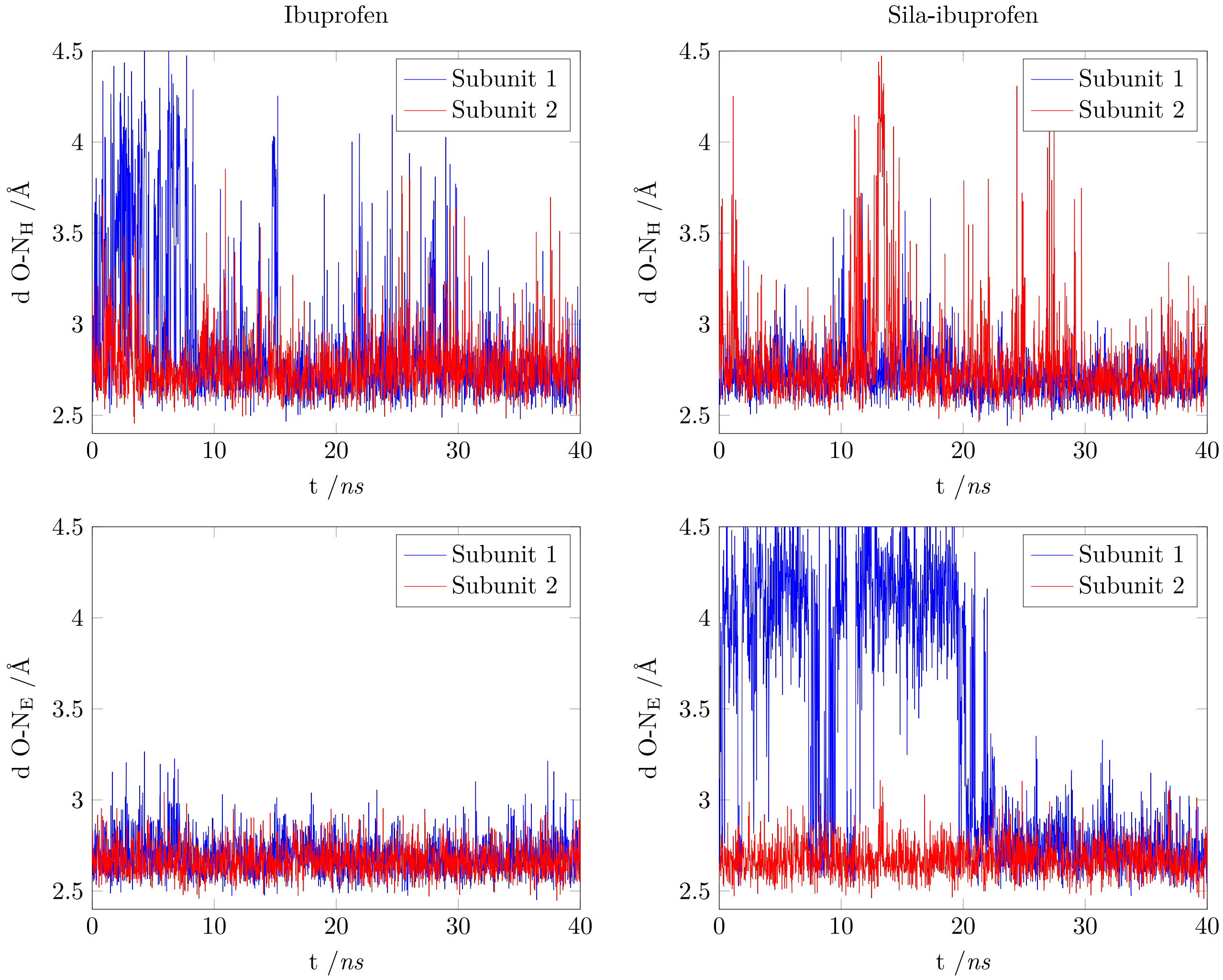


Figure 5.19: Plots of distances between ibuprofen (left) and sila-ibuprofen (right) oxygen atoms and arginine nitrogen atoms in **COX-I** during the production runs for both subunits of the enzyme. The atom labels for the nitrogen atoms are shown in Figure 5.18.

database. They are consistent between the simulations of ibuprofen and sila-ibuprofen, since they are based on identical input with only variation of atoms of ibuprofen into sila-ibuprofen.

One can clearly distinguish the interaction between N_H and N_E of arginine with ibuprofen and sila-ibuprofen in Figures 5.19 and 5.20. While the contact between the ibuprofen variants and arginine nitrogen N_H has a big variance in the region between 2.5 Å and 4 Å, the contact involving N_E is comparably stable and close throughout the simulations. The only outlier of this behavior is subunit 1 of sila-ibuprofen in COX-I during the first 20 ns of the production run. Here the distance is mainly around 4Å and only in the second half of the simulation binds as stable as subunit 2. If one investigates further, there is a second residue that might form hydrogen bonds with the carboxylic acid functional group of the drug molecule, which is a tyrosine residue, which is closer in COX-I than in COX-II during the simulations. Due to this behavior and the outlier in subunit 1 the analysis in the following sections was performed using subunit 2.

However, one can already deduce some insight into the behavior of both substances in the active site of the COX-enzymes: The position inside the enzyme seems to be very stable. No molecule, neither ibuprofen nor sila-ibuprofen, left the active site throughout the simulations. Only secondary interactions loosened up while at all times at least one of the anchoring bonds remained stable in a time frame of 40 ns, which is already a significant time frame analyzed for common docking studies. This leads to the conclusion that sila-ibuprofen could in principle have similar properties regarding binding affinity and stability to the enzymes and might

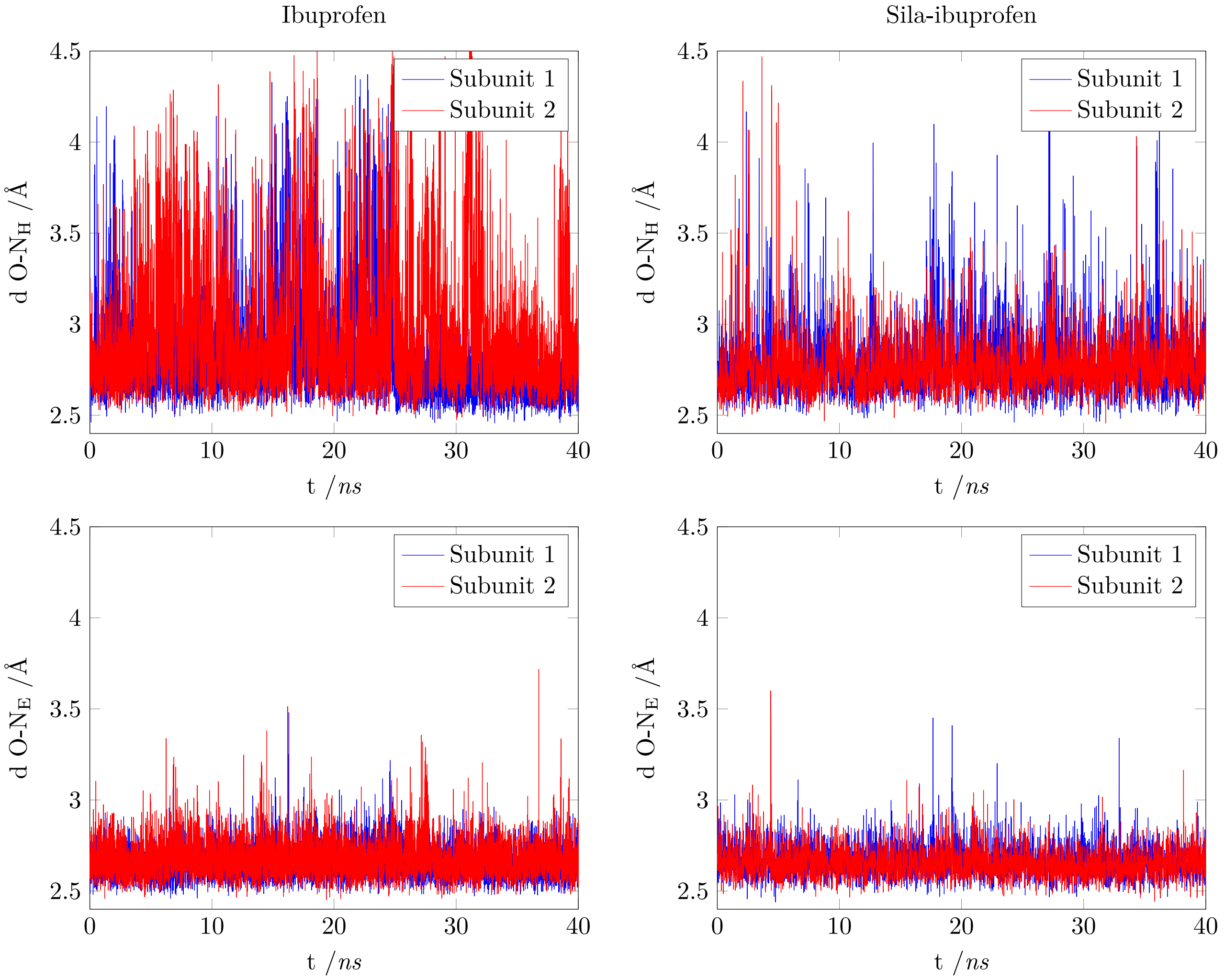


Figure 5.20: Plots of distances between ibuprofen (left) and sila-ibuprofen (right) oxygen atoms and arginine nitrogen atoms in **COX-II** during the production runs for both subunits of the enzyme. The atom labels for the nitrogen atoms are shown in Figure 5.18.

possibly be also active in terms of pain relief.

5.2.9 aNCI

Since the binding in the active site is reversible in both enzymes, it is of interest which specific contacts are most important and how they differ between the two drug substances. The NCI is a powerful descriptor for the visualization and understanding of non-covalent bonding. [183, 184] Since ibuprofen is a non-covalently bound inhibitor, this descriptor should provide good visual and qualitative insight into the binding of the drug in the active site. But since the NCI is highly dependent on the geometry at hand and since the thermal motion of the molecular dynamics make it difficult to distinguish between thermal contacts and permanently stable contacts, the aNCI, being the average over a series of geometries, was implemented as described in section 2.3.3 and used on the subunit 2 of both molecules, as discussed above. The final frame of the production runs was taken and subsequently a simulation of 1000 ps was started, where each ps a geometry was saved during the simulation. Atoms of ibuprofen or sila-ibuprofen were fixed, to allow later analysis of cubes around them more easily.
These geometries were subsequently used to calculate the aNCI based on promolecular densities using *cuQCrT*. In this case, the entities defined for the intermolecular plotting are the ibuprofen or sila-ibuprofen molecule as one entity and the protein as the second entity in the input file. Water molecules and salt ions were removed

from the geometry files for the analysis. A radius of 6 Å around the ligand molecule was chosen as size of the grid to be evaluated to accommodate as many interactions as possible. These calculations were performed on a desktop computer with a 4-core CPU and 2 consumer GPUs (NVIDIA-GeForce 1080Ti) over night. The resulting cubes then only contain non-covalent interactions that are persistent over the whole simulation timespan and can therefore be considered thermally stable. The plots of these files for COX-I are shown in Figure 5.21 and for COX-II in Figure 5.22. The calculation of the properties was performed taking into account the ca. 157000 atoms of the complete dimer of protein strands of the corresponding COX.

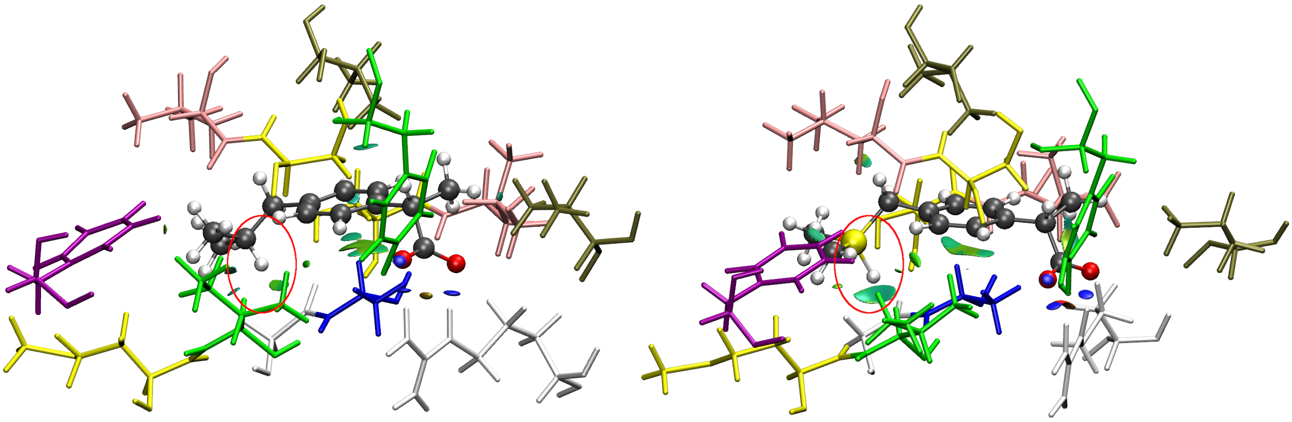


Figure 5.21: Plot of isosurfaces of the aNCI between ibuprofen (left) and sila-ibuprofen (right) and the active site of **COX-I**. All atoms of the enzyme were used for the calculation, but only closest residues visualized. Color coding of residues: Arg & Ala = white, Gly = blue, Met & Ser = yellow, Tyr & Ile = green, Phe = purple, Val = brown, Leu = pink. Not all residues shown for clarity of the picture. Visualized using VMD.

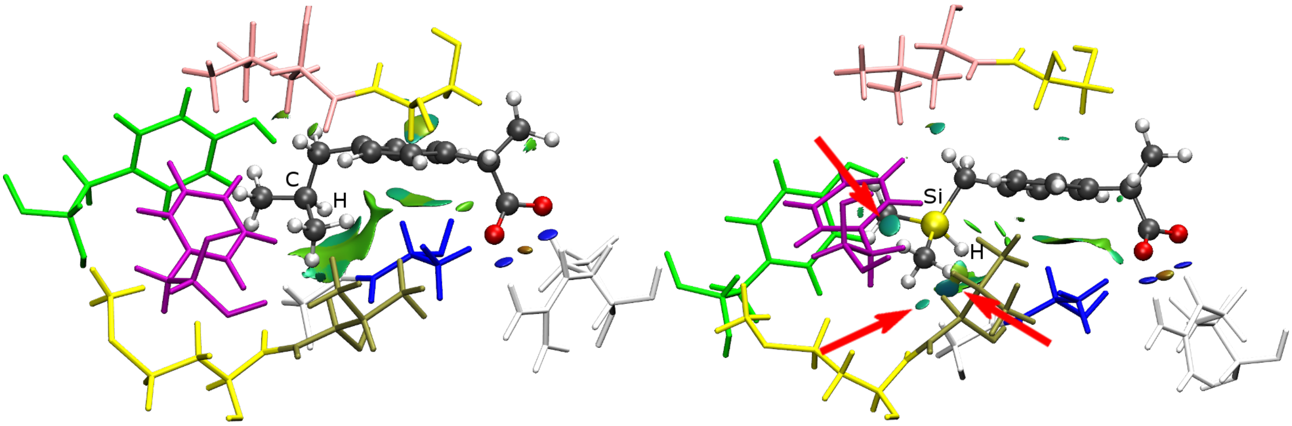


Figure 5.22: Plot of isosurfaces of the aNCI between ibuprofen (left) and sila-ibuprofen (right) and the active site of **COX-II**. All atoms of the enzyme were used for the calculation, but only closest residues visualized. Color coding of residues: Arg & Ala = white, Gly = blue, Met & Ser = yellow, Tyr = green, Phe = purple, Val = brown, Leu = pink. Not all residues shown for clarity of the picture. Visualized using VMD.

The interactions in COX-I show similar interactions of dispersive kind (green colored isosurface of the aNCI) for the interaction of the benzyl group of ibuprofen and sila-ibuprofen in both cases. They interact in both cases with the neighboring gly-ala residues of the active site. The carboxylic acid function on the right side of the molecule is involved in strong hydrogen bond interactions with the arginine residue, as expected, visualized by blue surfaces for the bonds and an orange surface in between for the ring strain.

A difference observable in Figure 5.21 is the polarity, since the *umpolung* of the Si-H bond when compared to the usual C-H bond yields different electrostatic behavior of the molecule (compare section 1.7). This leads to a significantly stronger interaction around this region, highlighted by the red circles.

A major difference is the distribution of surfaces around the methyl groups of the isobutyl group in ibuprofen and the dimethylsilyl group in sila-ibuprofen highlighted with red arrows. Here, a more pronounced surface can be seen around the methyl groups of sila-ibuprofen when compared to Ibuprofen. This might be rationalized by a better fit inside the pocket, as the target molecule of COX enzymes is arachidonic acid, which is a bent all-*cis* fatty acid, that requires some space in the active site. Since ibuprofen is a branched molecule with the isobutyl function it blocks the space in the hydrophobic region of the active site. The difference in size of carbon and silicon make the same function in sila-ibuprofen more space demanding, since the distance between silicon and the carbon atoms of the methyl groups is higher. This way the molecule can occupy more space and have more surface interaction with the active site, while retaining the structural motif and shape of ibuprofen.

This raises the question what nature the interactions with the residues around the silane function have and whether there is a significant difference in the binding strength. It can be concluded however, that the major interactions of both molecules with the respective active sites of COX-I and COX-II are similar in directions and size, except for the silane function, which could hint towards a similar behavior in terms of their biological activity and potential as COX inhibitors.

5.2.10 Averaged Interaction Energies (aIE)

Since the aNCI is only qualitative with no link to energies, a different analysis is necessary. The averaged interaction energies were calculated, using the aIE script as introduced in section 3.2. The wavefunctions on a level of theory of B3LYP/6-31G(d,p) of all required residues, saturated with hydrogen atoms if bonds were cut, were calculated using *Gaussian09* for geometries taken from a trajectory file obtained after MD. Subsequently the Interaction Energy was calculated using *Tonto* (compare section 2.3.6). Residues of interest were selected from the dynamics simulation based on the criterion that they were not water or salt ions and had at least one atom within 3 Å of the ibuprofen or sila-ibuprofen molecule. The same trajectory as prepared for the aNCI was used for these calculations.

Although there is significant fluctuations in the interaction energy contributions between all amino acids of COX-I and ibuprofen in Figure 5.23, the values do not show any trends. The magnitude of fluctuations is comparable to that of distances in the active site, as seen in Figures 5.19 and 5.20. Five residues are outstanding in these simulations: Arg, Tyr, GlyAla, MetIle and LeuSer. Arg and Tyr dominate the electrostatic and polarization plots. This confirms the observed strong hydrogen bond and the pronounced contact of the tertiary hydrogen atom in ibuprofen and the silane hydrogen atom in sila-ibuoprofen in the aNCI plots and the literature referring to the arginine-ibuprofen contact as the anchor of ibuprofen inside the active site. [299, 301] The correlation between electrostatic and polarization energies can be understood taking into account the fact that a strong electrostatic interaction, like hydrogen bonds, will in most cases also lead to a polarization of the whole molecule. The residues GlyAla, MetIle and LeuSer are the residues surrounding the hydrophobic part of ibuprofen when bound in the active site. These residues exhibit mainly dispersive interactions with the respective drug molecule. These are the highest contributors in the dispersion plots of Figures 5.23 and 5.25.

The biggest difference between the interaction energies of ibuprofen and sila-ibuprofen is observed for the residue MetIle, which shows a higher dispersion interaction and also slightly higher polarization energies in sila-ibuprofen compared to ibuprofen, while GlyAla has slightly lower dispersive interaction energies in sila-ibuprofen. The different contributions of the four energy terms can be summed into an overall energy of interaction between two molecules employing scale factors, as mentioned in section 2.3.6. For the used level of theory the scale factor for the total energy is

$$E_{Tot} = 1.057E_{ele} + 0.74E_{pol} + 0.871E_{dis} + 0.618E_{rep}. \tag{5.1}$$

Using this equation the total energy was plotted for the whole simulation in COX-I in Figure 5.24.

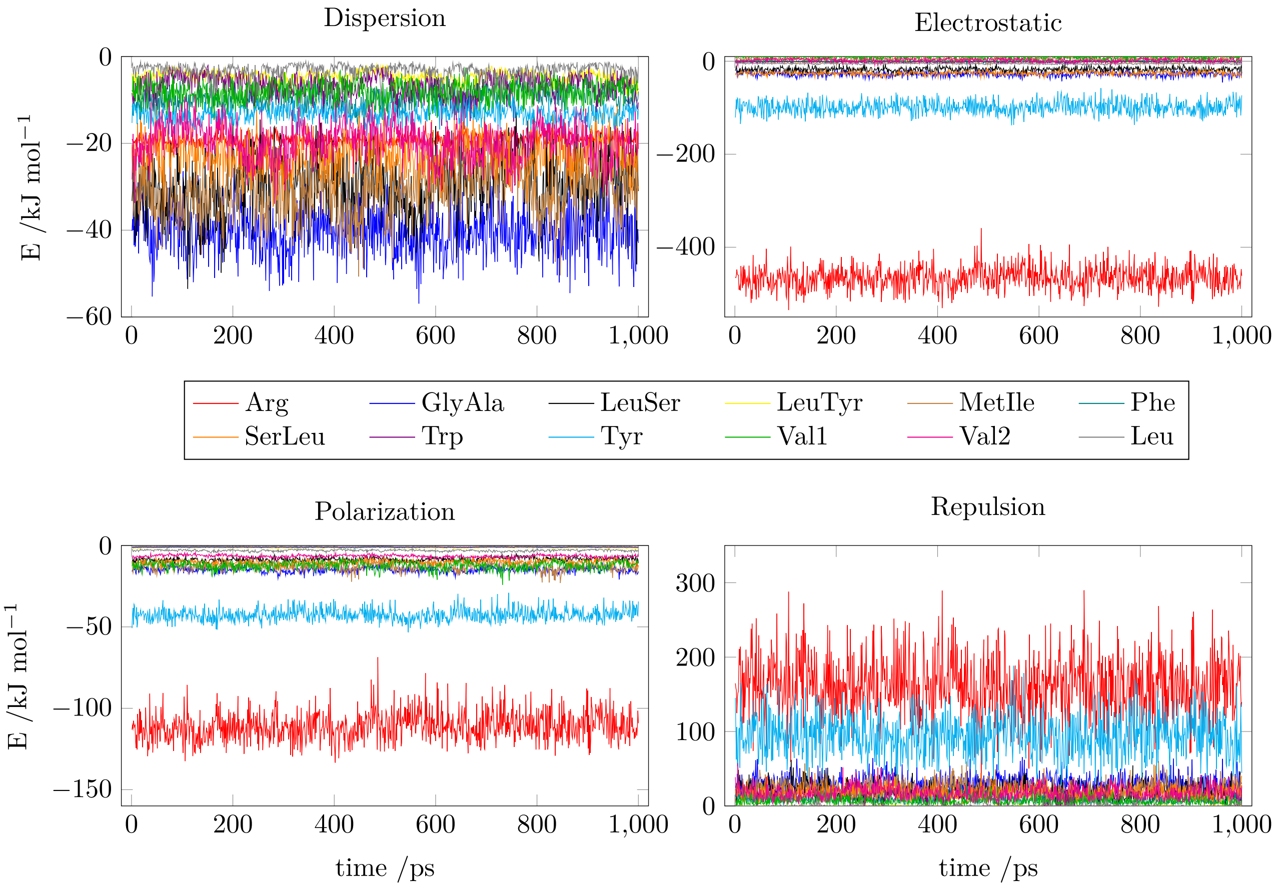


Figure 5.23: Plots of interaction energies of **ibuprofen** and amino acid residues of **COX-I** within a radius of 4 Å around ibuprofen as implemented in *CrystalExplorer*, calculated using the procedure described in section 3.2. From top left to bottom right: dispersion energy, electrostatic energy, polarization energy and repulsion energy against time during the simulation.

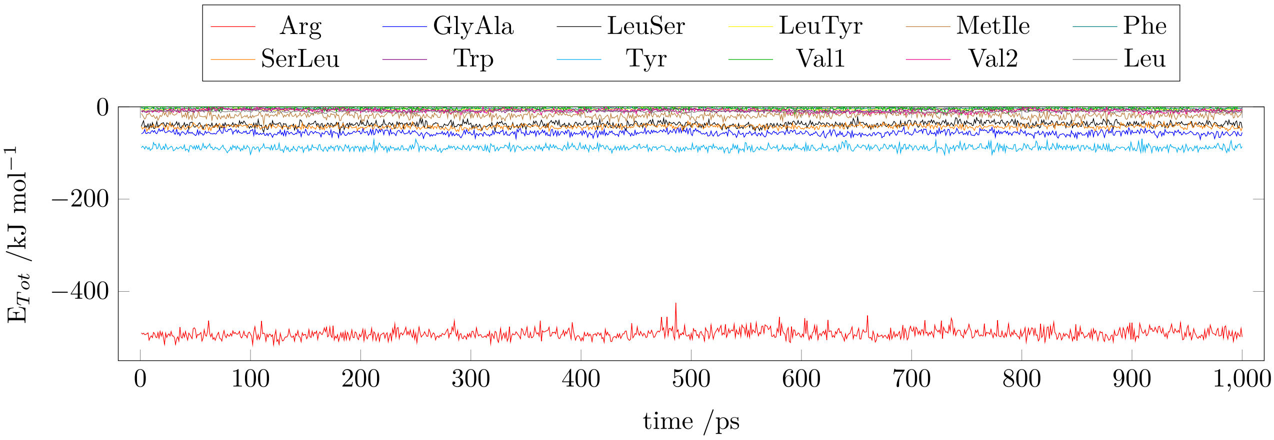


Figure 5.24: Plot of the total interaction energy of the simulation of **ibuprofen** and amino acid residues of **COX-I** within a radius of 3 Å around ibuprofen. Calculated from values plotted in Figure 5.23 using equation 5.1.

The same plots are shown for sila-ibuprofen in Figures 5.25 and 5.26.

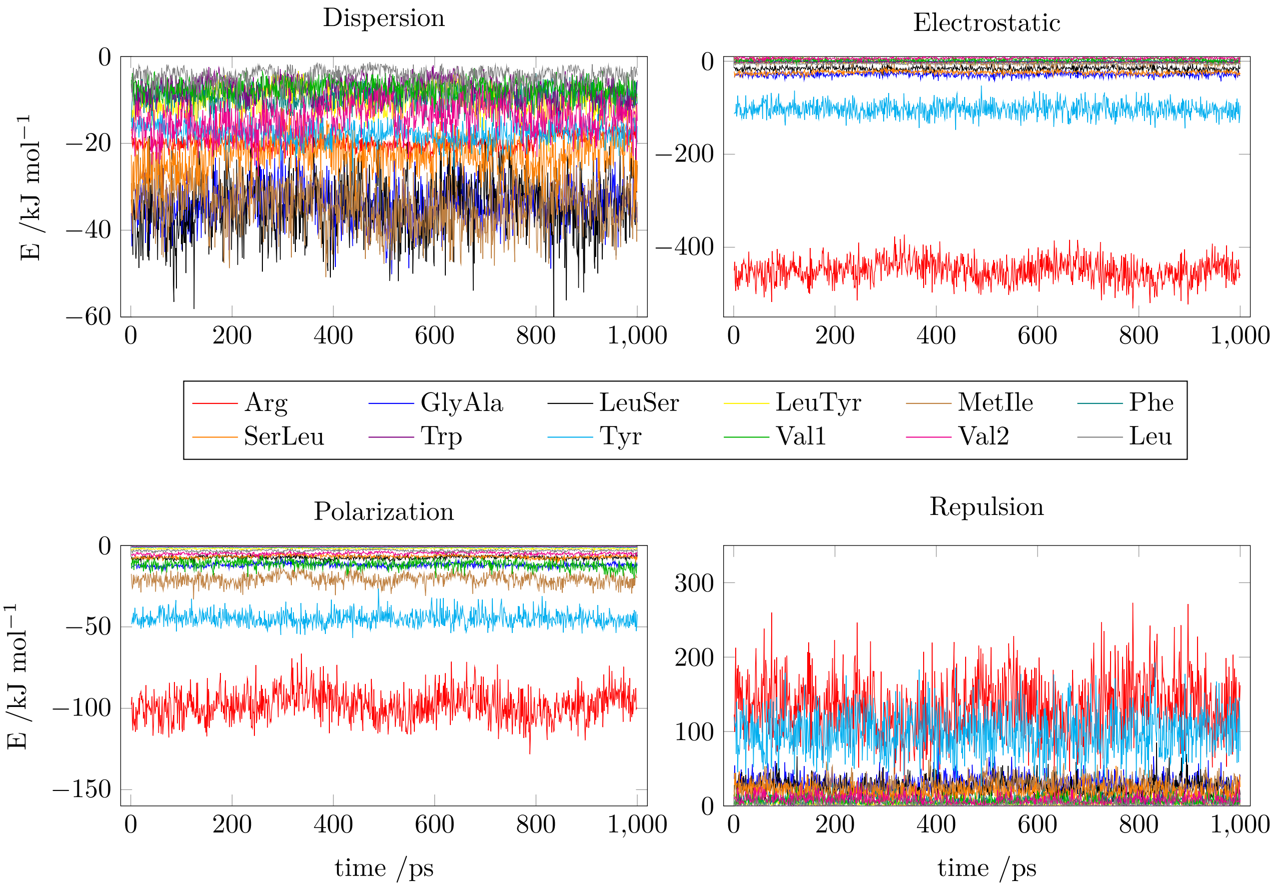


Figure 5.25: Plots of interaction energies of **sila-ibuprofen** and amino acid residues of **COX-I** within a radius of 3 Å around sila-ibuprofen as implemented in *CrystalExplorer*, calculated using the procedure described in section 3.2. From top left to bottom right: dispersion energy, electrostatic energy, polarization energy and repulsion energy against time during the simulation.

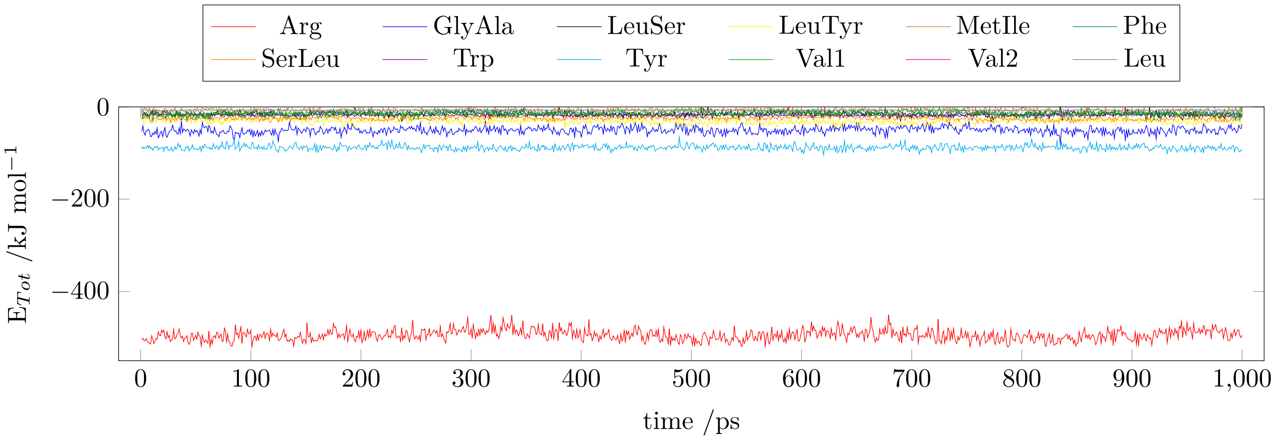


Figure 5.26: Plot of the total interaction energy of the simulation of **sila-ibuprofen** and amino acid residues of **COX-I** within a radius of 4 Å around sila-ibuprofen. Calculated from values plotted in Figure 5.25 using Equation 5.1.

To get a better understanding of the individual contributions of residues and the thermal fluctuation during the simulation for both drug molecules bound to COX-I the averages and standard deviations of all values are reported in Table 5.6.

The values are similar between both derivatives of ibuprofen. Magnitude and sign of energies are very close in

Table 5.6: Table of arithmetic averaged interaction energies of ibuprofen and sila-ibuprofen in **COX-I** in kJ mol^{-1}. Numbers in brackets denote standard deviation over the 1000 frames which were averaged.

Residue	Ibuprofen					Sila-buprofen				
	$\langle E_{Ele} \rangle$	$\langle E_{Disp} \rangle$	$\langle E_{Pol} \rangle$	$\langle E_{Rep} \rangle$	$\langle E_{Tot} \rangle$	$\langle E_{Ele} \rangle$	$\langle E_{Disp} \rangle$	$\langle E_{Pol} \rangle$	$\langle E_{Rep} \rangle$	$\langle E_{Tot} \rangle$
Arg	-449(25)	-19.7(16)	-98(9)	134(39)	-494(12)	-465(24)	-19.1(11)	-111(8)	161(39)	-492(10)
GlyAla	-29(4)	-34(4)	-12.0(10)	31(10)	-51(7)	-28(5)	-40(5)	-14.9(15)	30(10)	-57(4)
LeuSer	-16(4)	-35(7)	-7.4(9)	25(11)	-16(5)	-18(4)	-29(7)	-8.6(12)	20(10)	-38(5)
LeuTyr	-1.7(14)	-10(3)	-1.9(4)	4(3)	-31(5)	-0.6(11)	-4.6(12)	-1.34(16)	0.4(5)	-5.4(16)
Leu	-5.4(10)	-4.5(19)	-3.2(5)	2(2)	-7.1(9)	-5.3(9)	-3.3(15)	-3.1(5)	0.7(14)	-10.2(17)
MetIle	-2(5)	-34(6)	-21(3)	27(10)	-9(5)	5(4)	-31(6)	-13(2)	22(9)	-18(5)
Phe	0.1(14)	-9(2)	-1.09(12)	8(5)	-15(4)	-0.1(14)	-10(2)	-1.13(13)	10(5)	-3.3(9)
SerLeu	-24(3)	-23(4)	-6.9(9)	19(8)	-25(4)	-26(3)	-23(4)	-10.3(13)	19(8)	-43(4)
Trp	-1.8(8)	-7(2)	-0.57(7)	0.5(26)	-17(2)	-1.7(8)	-7(2)	-0.74(10)	1(3)	-8(2)
Tyr	-103(15)	-18(2)	-45(4)	98(29)	-88(5)	-99(13)	-13.2(17)	-43(4)	95(27)	-89(5)
Val1	3(3)	-7.3(19)	-12(2)	6(4)	-14(4)	10(4)	-8(2)	-12(2)	7(4)	-1(4)
Val2	6.3(20)	-14(3)	-4.9(7)	8(5)	4(2)	2(3)	-20(4)	-6.5(8)	18(8)	-9(3)
Σ	-622(65)	-216(40)	-213(23)	364(128)	-762(56)	-626(65)	-207(38)	-225(22)	383(125)	-773(47)

both calculations. This means there is the same kind of forces acting on the molecules in the active site which should yield similar binding strength of sila-ibuprofen for COX-I as for ibuprofen, when using the energetics to make an assumption on binding affinities. Differences in polarization and dispersion terms of the interaction energies are mostly present in the vicinity of the sila-substituted position which coincides with the areas where an increase of the aNCI was observable in the circled regions when introducing the silicon atom in Figure 5.21. Since the coloured surfaces in the aNCI plots mainly depict dispersion, which means low electron density, an increase of the surface corresponds to more density being present in that region. Since the negatively charged hydride in sila-ibuprofen (compare Figure 5.12) has a higher electron density this might lead to higher electron densities to form these interactions. The polarization terms are based on two factors: the polarizability and the electric field at the atomic position. In the case of ibuprofen the charges are relatively small within the molecule itself, therefore it might be expected that the polarization interaction with the environment is relatively small. Sila-ibuprofen on the other hand has pronounced charges and a intramolecular dipole (compare section 5.2.5) which can polarize the pocket and therefore lead to polarization interaction even if there is mainly hydrophobic amino acids present in the vicinity. This effect is reflected by the mostly higher polarization terms in the case of sila-ibuprofen, especially near the silane function.

Although the standard deviation of the energies seems relatively high in some cases (50 kJ mol^{-1} for the total energy) the relative trend when going from ibuprofen to sila-ibuprofen is qualitatively understandable and the total energies are in a reasonable scale, which means this method can describe the binding of a molecule within the pocket correctly. The high deviation of the energies is due to the thermal motion and not due to inaccuracies within the calculation of the energies themselves. This can be nicely correlated with the repulsive energy terms having much higher deviations compared to the other energy contributions, as the repulsion is due to the residues coming closer due to thermal motion and being repelled which leads to a decrease of the energy for this residue in the later steps. Therefore a simulation of biological systems at biologically relevant temperatures will most likely always show this order of deviations around the average energy for all energetic contributions. Only a much stronger binding force being present might reduce these numbers, for example by multiple functional groups being able to bond in the enzyme pocket and therefore reducing the flexibility of the molecule to move within the pocket.

Since the thermal movement of atoms is included in these calculations, the method of averaged interaction energies could yield much more accurate interaction energies between an active site of a protein and ligand. However, this method would work best and most convenient if interfaced to a QM/MM MD where the wavefunctions are readily available. This kind of calculation might then be included easily if two layers of QM-regions are defined and the wavefunctions transferred and analyzed as in this case. This could also yield new methods for docking studies, where the geometry is then generated, the wavefunctions calculated and energies calculated automatically and the results used for better determination of docking scores, without the need for force-field biased docking scores. The question whether the fluctuation of these energies might be interpreted as a measure for entropy and maybe even converted into them is beyond the scope of this thesis.

The same type of calculations was performed for the sila-/ibuprofen complex with COX-II. The results will be presented in the same manner and order as for COX-I in Figures 5.27 to 5.30 and Table 5.7. Fortunately, the vicinity of the ibuprofen molecule in the active site of COX-II is very similar to that in COX-I. Only one mutation is found in COX-II where MetIle is MetVal compared to COX-I, which is also reported in the literature. [478] This difference, considering that the chemical deviation between Ile and Val is only one CH_2 group, should only marginally change the shape of the pocket and allows nice side-by-side comparison for the per residue energetic contributions.

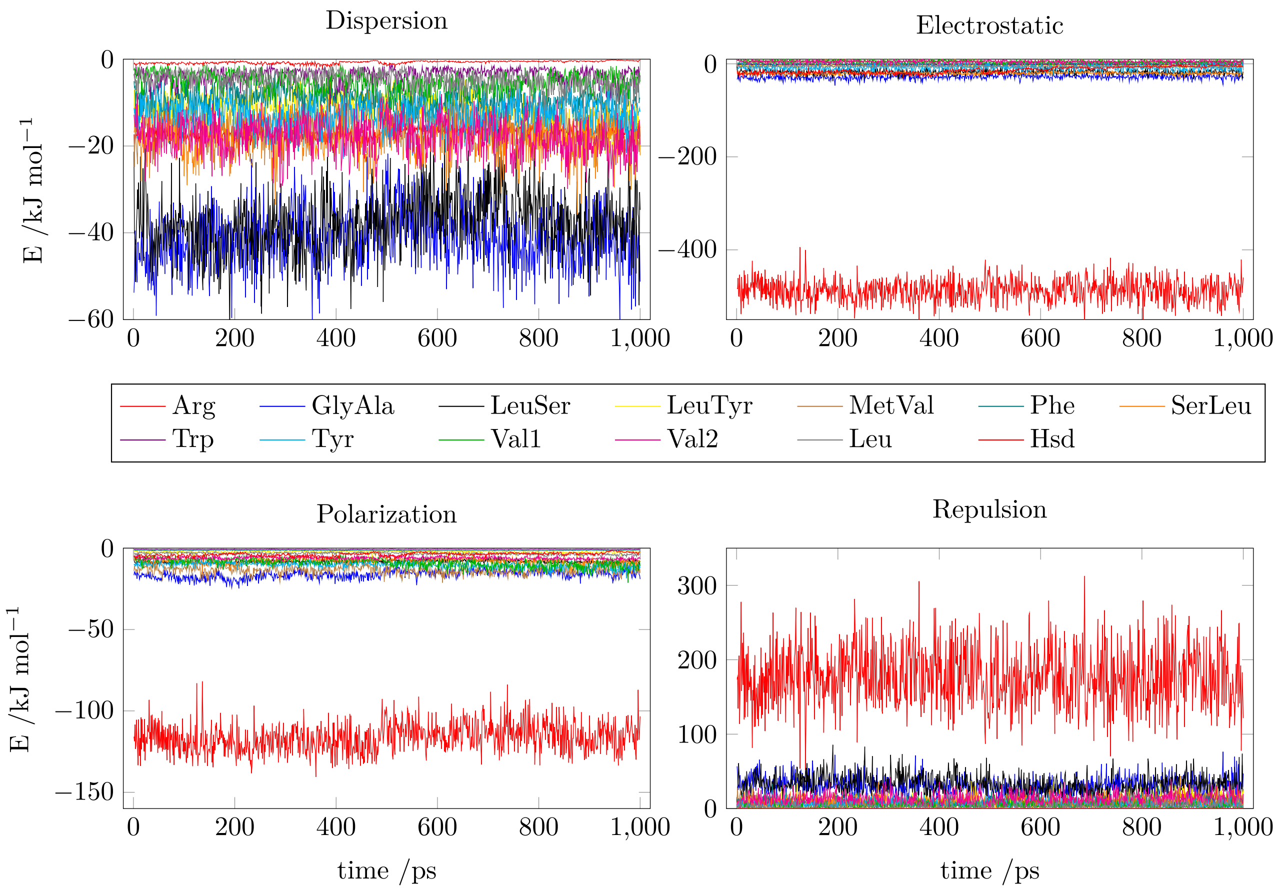


Figure 5.27: Plots of interaction energies of **ibuprofen** and amino acid residues of **COX-II** within a radius of 3 Å around ibuprofen as implemented in *CrystalExplorer*, calculated using the procedure described in section 3.2. From top left to bottom right: dispersion energy, electrostatic energy, polarization energy and repulsion energy against time during the simulation.

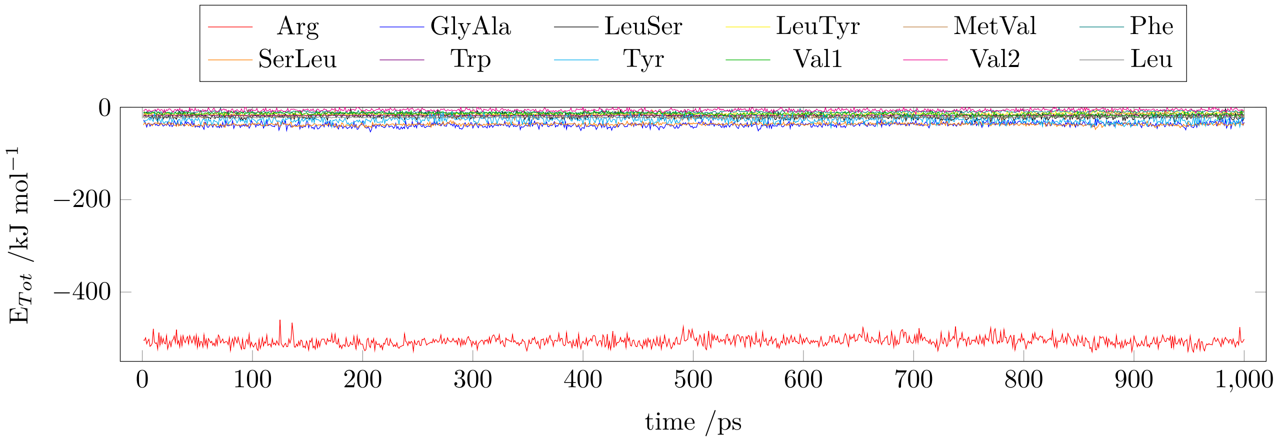


Figure 5.28: Plot of the total interaction energy of the simulation of **ibuprofen** and amino acid residues of **COX-II** within a radius of 4 Å around ibuprofen. Calculated from values plotted in Figure 5.27 using equation 5.1.

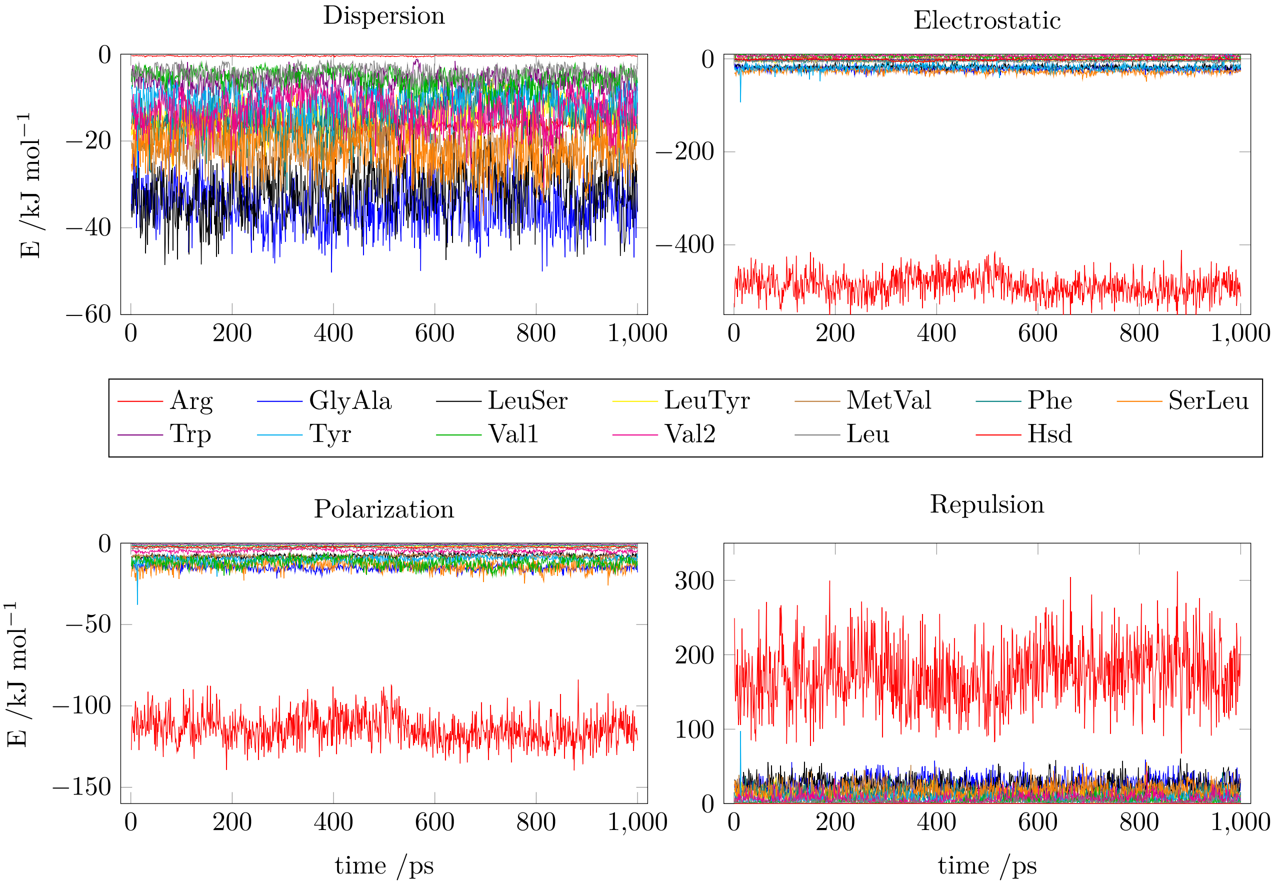


Figure 5.29: Plots of interaction energies of **sila-ibuprofen** and amino acid residues of **COX-II** within a radius of 3 Å around sila-ibuprofen as implemented in *CrystalExplorer*, calculated using the procedure described in section 3.2. From top left to bottom right: dispersion energy, electrostatic energy, polarization energy and repulsion energy against time during the simulation.

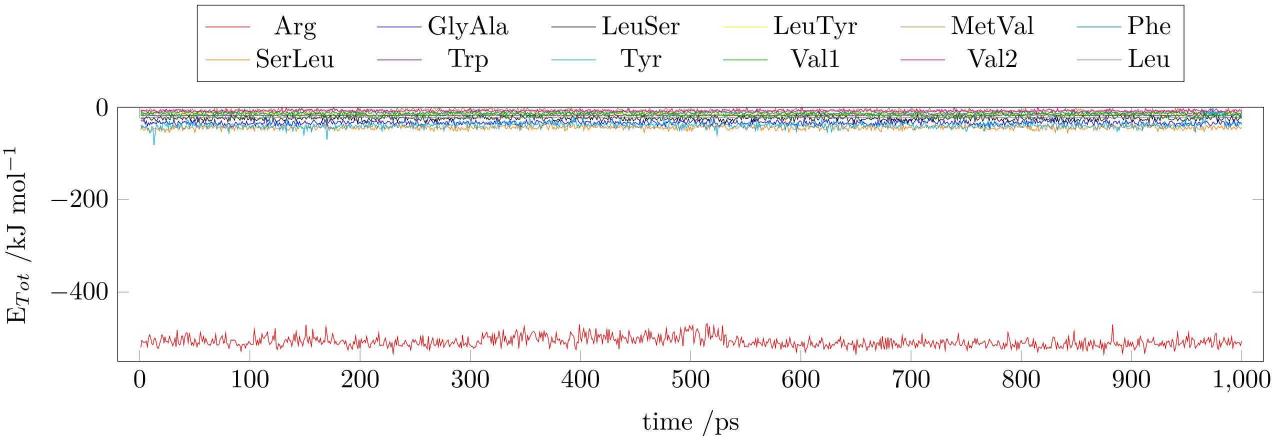


Figure 5.30: Plot of the total interaction energy of the simulation of **sila-ibuprofen** and amino acid residues of **COX-II** within a radius of 4 Å around sila-ibuprofen. Calculated from values plotted in Figure 5.29 using equation 5.1.

Table 5.7: Table of arithmetic averaged interaction energies of ibuprofen and sila-ibuprofen in **COX-II** in kJ mol^{-1}. Numbers in brackets denote standard deviation over the 1000 frames which were averaged.

	Ibuprofen					Sila-buprofen				
Residue	<E_{Ele}>	<E_{Disp}>	<E_{Pol}>	<E_{Rep}>	<E_{Tot}>	<E_{Ele}>	<E_{Disp}>	<E_{Pol}>	<E_{Rep}>	<E_{Tot}>
Arg	-487(24)	-17.0(10)	-116(9)	178(40)	-508(9)	-487(25)	-16.3(9)	-114(9)	173(41)	-508(11)
GlyAla	-28(4)	-41(6)	-16(2)	31.0(11)	-38(5)	-22(4)	-35(5)	-15.1(15)	26(9)	-34(4)
Hsd	-14(7)	-0.7(3)	-3.8(10)	0(0)	-32(8)	-1.6(18)	-0.41(8)	-2.4(3)	0(0)	-19(2)
LeuSer	-19(5)	-38(7)	-8.2(9)	33(13)	-20(5)	-18(4)	-31(6)	-7.7(11)	25(1)	-24(4)
LeuTyr	-3(2)	-12(3)	-2.4(6)	9(6)	-15(2)	-2.9(17)	-14(3)	-1.7(3)	9(6)	-14(2)
Leu	-5.8(13)	-7(2)	-3.3(6)	3(3)	-21.7(10)	-5.4(11)	-4.4(17)	-3.1(5)	2(2)	-21.9(8)
MetVal	2(4)	-15(3)	-14(2)	8(5)	-18(4)	2(3)	-21(4)	-8.3(10)	18(8)	-8(3)
Phe	0.8(14)	-9(2)	-1.36(15)	8(5)	-10.2(19)	-3(2)	-14(4)	-1.5(3)	12(7)	-12(2)
SerLeu	-20(2)	-18(4)	-8.1(16)	9(6)	-37(3)	-27(4)	-22(5)	-14(3)	16(8)	-44(4)
Trp	-1.9(7)	-3.7(15)	-0.77(12)	0.1(12)	-17.6(8)	-1.7(7)	-6(2)	-0.71(11)	0.1(12)	-17.4(8)
Tyr	-8(5)	-13(3)	-10.4(17)	6(4)	-28(6)	-19(7)	-11(3)	-9.9(18)	6(5)	-39(7)
Val1	5(2)	-5(2)	-9(2)	3(3)	-14(3)	5(3)	-5(2)	-12(3)	3(3)	-16(3)
Val2	5(2)	-18(4)	-6.0(8)	13(7)	-6(3)	6(2)	-13(4)	-4.7(8)	8(5)	-8(2)
Hsd	-14(7)	-0.7(3)	-3.8(10)	0(0)	-32(8)	-1.7(18)	-0.41(8)	-2.4(4)	0(0)	-19(2)
Σ	-574(90)	-195(40)	-199(23)	301(103)	-764(50)	-574 (60)	-195(40)	-195(22)	296(105)	-765(45)

In comparison between COX-I and COX-II, it is remarkable that it seems like the Tyr residue does not play a significant a role in the binding of both molecules inside COX-II when compared to the situation in COX-I. This is based on the significantly decreased electrostatic interaction energy in comparison to COX-I while the energy of the interaction with Arg is very similar. Comparing the plots of hydrogen bond lengths between Arg and ibuprofen/sila-ibuprofen in Figures 5.19 and 5.20 shows that the N_H residue is more loosely bound in COX-II which coincides with the lower binding energy with Tyr, while within the thermal fluctuations the contact to Arg is more or less similar, even slightly stronger bonded. The side of N_H of the twofold hydrogen bond is that one which is closer to the Tyr residue in both active sites. Maybe this behavior is due to the fact that, since no other partner for interactions is present around the carboxylic acid, the twofold hydrogen bond is more unperturbed and therefore stronger. It appears that the supporting role that Tyr has in COX-I keeps the second hydrogen bond in closer proximity, while this additional support is missing in COX-II, but does not directly affect the interaction energy as a whole. It has to be kept in mind that the simulations for the aIE and aNCI are different simulations than the ones in the distances plots 5.19 and 5.20. The aNCI and aIE are performed after those simulations with different trajectory spacing.

In fact, the total energy of residues interacting with ibuprofen or sila-ibuprofen is similar in both enzymes COX-I and COX-II, taking into account the uncertainty of the values due to thermal fluctuation. In COX-I the total energy of ibuprofen is -762 kJ mol^{-1}, in COX-II it is -764 kJ mol^{-1}. The interaction strength is almost identical between the two envrionments. In the case of sila-ibuprofen the energies are -773 kJ mol^{-1} and -765 kJ mol^{-1}, respectively. In both cases the uncertainties are around 50 kJ mol^{-1}. This means within thermal fluctuation the binding strength between the ligands and the enzyme pocket is identical among all ligands and the two modifications of COX. Based on these similarities it is assumed that ibuprofen and sila-ibuprofen are in fact bioisoster and sila-ibuprofen should at least exhibit similar medicinal potency. [240, 292]

These energies correspond to the interaction energy keeping the ligand within the active site, so from an energetic point of view the two substances should have similar forces acting upon them to keep them in the pocket and therefore have similar time scales for the time the molecule resides within the active site and therefore inhibits the enzyme. However, the process of inhibition is not only driven by the energetic contribution once the ligand is inside the pocket but also by kinetic effects and energetics associated with the structural change upon the docking of an inhibitor molecule. These will be adressed in the next section 5.2.11.

5.2.11 Free Energy Perturbation Calculations

FEP simulations were performed in order to assess the difference in the complete binding energy between drug and enzyme taking into account the energetic contribution from the reshaping of the active site in addition to interaction energies between the protein and the ligand. These simulations were performed using *NAMD2* and similar settings as for the previous MD (compare section 5.2.7). Annealing and equilibration runs were performed where $\lambda = 0$, which was chosen to be the ibuprofen conformation. The production run was performed using λ steps of 0.025 with 600,000 steps and 100,000 steps for *alchEquilSteps*. Simulations were run in forward and backwards mode, meaning a incremental run starting from $\lambda = 0$ to 1 and decremental run from $\lambda = 1$ to 0. The switching was chosen to have a *vdwlambdaEnd* parameter of $\lambda = 1$, the electrostatic switch was set to take place at $\lambda = 0.5$, the switching coefficient was chosen to be 5. A structure file for the generation of *.psf* information was created, where the merged bonds and angles of ibuprofen and sila-ibuprofen are defined. A visualization of this situation can be seen in Figure 5.31.

The close overlap of atoms is possible because the FEP algorithm crosses out energy contributions arising from those atoms that are being morphed into each other, which are shown with different colours in Figure 5.31. The force field is then read from a merged file, that contains all necessary information. The result of the transformation simulations for COX-I is shown in Figure 5.32.

The free energy change in COX-I is estimated as 1.46(14) kJ mol^{-1} over the whole λ range by the FEP-analysis tool in *VMD*. The change in free energy is basically the difference between the left and right side of the plots. If the morphing of ibuprofen ($\lambda = 0$) into sila-ibuprofen ($\lambda = 1$) would lead to a change in free energy the plot would yield asymmetrical behavior and the right hand-side would either be significantly higher or lower. It is usual to perform the transformation starting from $\lambda = 0$ and increase stepwise (forward) and to decrease starting from $\lambda = 1$ (backward), to have a consistency check, as there might be artifacts introduced when e.g. switching between electrostatics at the midpoint of morphing. To assess the plausibility and consistency, both simulations are evaluated and the agreement between them used as a measure for consistency. As seen in Figure 5.32 the plots agree very nicely and only show very little deviation. Each of these points in the plot summarizes 600.000 time-steps of molecular dynamics, as mentioned in section 2.4.2. This big number of steps ensures consistency and reproducibility. If too few time steps were chosen the result might not be representative of the energy change

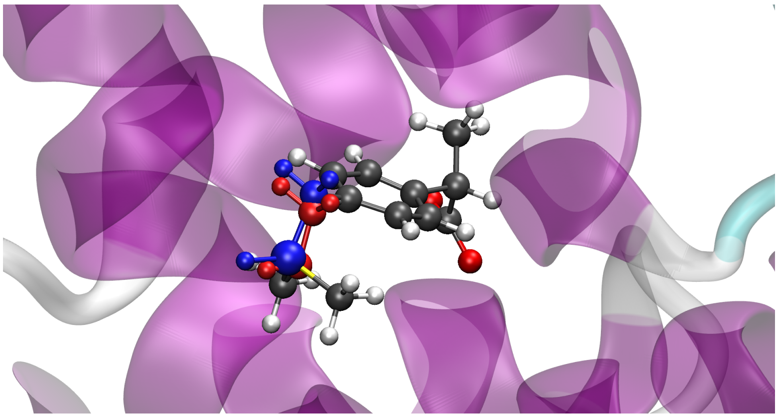


Figure 5.31: Visualization of residue created for FEP simulations in the vicinity of the protein during the simulation. Atoms color coded in blue are atoms of sila-ibuprofen, red atoms are ibuprofen. Background protein in *NewCartoon* style is COX-II, where ibuprofen/sila-ibuprofen is bound in the active site and morphed. Visualized using VMD. [56]

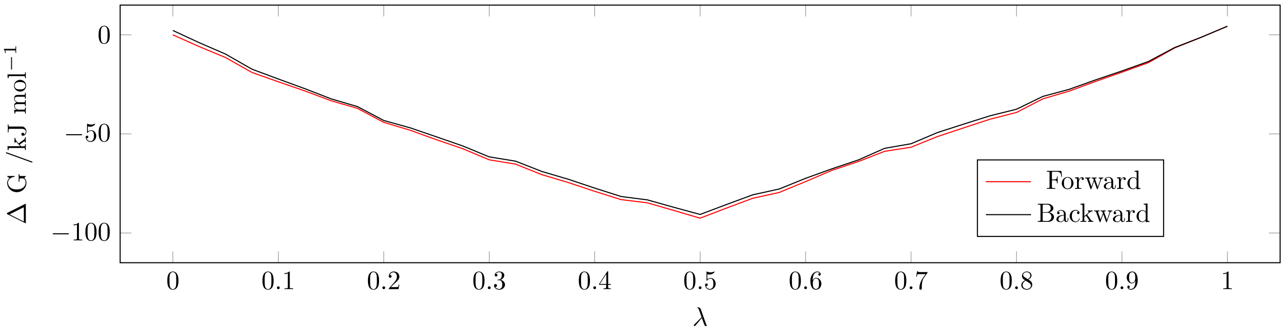


Figure 5.32: Plots of Free energy difference for *in-silico* alchemical transformation of ibuprofen (λ = 0) into sila-ibuprofen (λ = 1) inside the active site of COX-I.

and include fluctuation artifacts, which would lead to e.g. jumps in the Free Energy change or discrepancies between forward and backward run.

The small difference in free energy between ibuprofen and sila-ibuprofen indicates that both have equally strong affinity to the binding pocket, which is in agreement with the interaction energies and aNCI calculations.

The FEP simulations were also performed for the COX-II complex. The free energy plots are shown in Figure 5.33. Also inside COX-II the FEP calculation yields only very small differences. The change is estimated as -4.7(2) kJ mol^{-1} over the whole range of the simulation. This means there is close to no difference in the binding, which is in agreement with the observed similarities over all residues of the averaged interaction energies. (compare Tables 5.6 and 5.7)

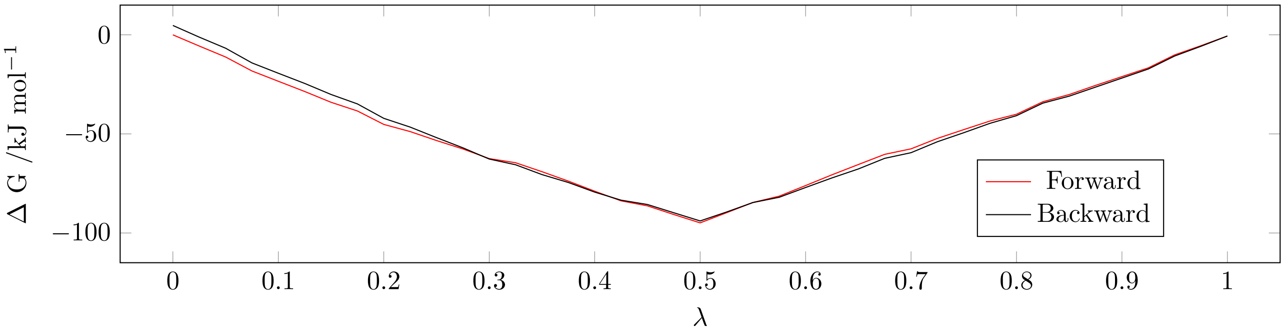


Figure 5.33: Plots of Free energy difference for *in-silico* alchemical transformation of ibuprofen (λ = 0) into sila-ibuprofen (λ = 1) inside the active site of COX-II.

These findings suggest that from all simulations and models a very high similarity in terms of binding affinity and energy is expected for ibuprofen and sila-ibuprofen, which should result in very similar inhibitory properties of both substances rendering them promising biosisosteres.

5.2.12 Measurement of IC $_{50}$

From all theoretical and model calculations sila-ibuprofen is a promising candidate for potentially at least identical medical properties as ibuprofen. Therefore measurements of the inhibitory potential of both ibuprofen and sila-ibuprofen were performed at *Eurofins CEREP* to have experimental evidence of possible similarities and differences of both substances. The measurement of the IC_{50} *in-vitro* for human COX-I and COX-II was performed in semi-logarithmic concentration steps from 0.1 μmol/l to 100 μmol/l and the results fitted by commercial software (for details see section D.2). All concentrations were measured in double determination and the average with standard deviations as error bars, as well as fitted models plotted in Figure 5.34. The fitting was performed using a model of

$$Y = D + \left(\frac{A - D}{1 + \left(\frac{C}{IC_{50}} \right)^{n}} \right). \tag{5.2}$$

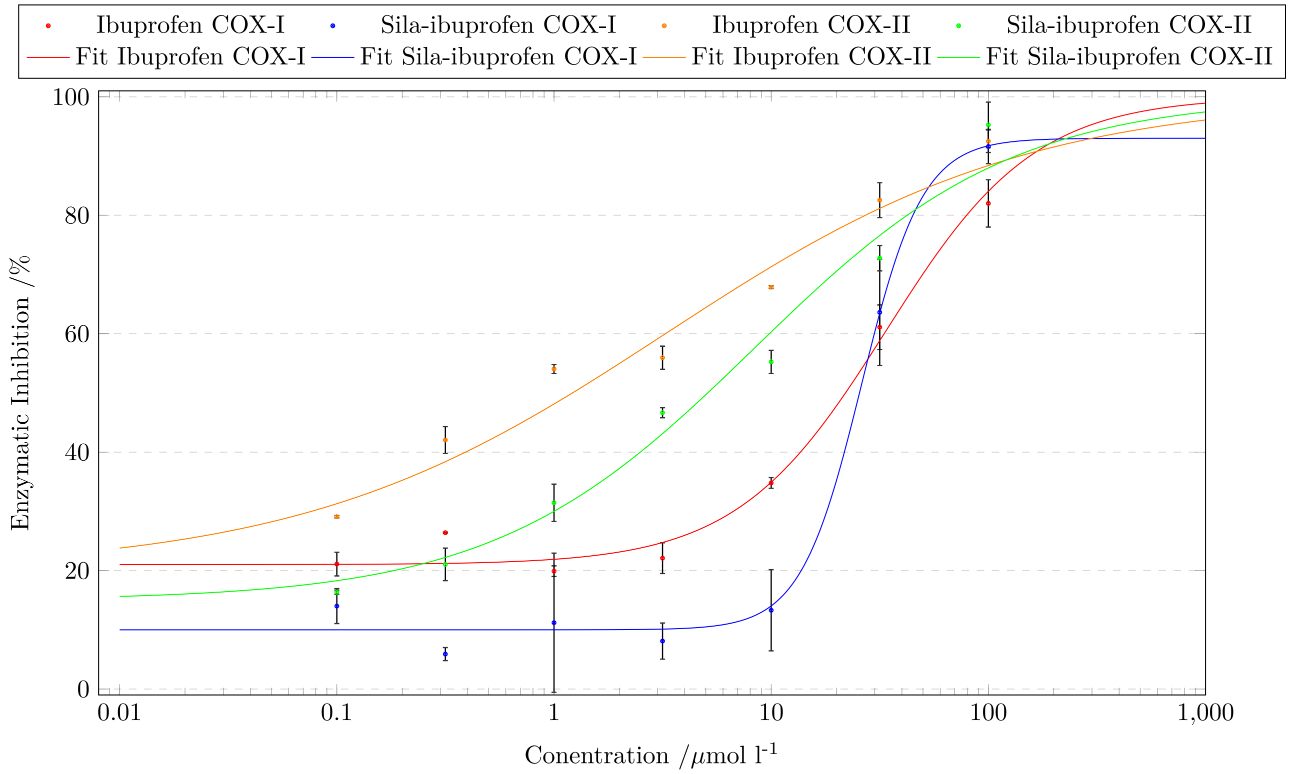


Figure 5.34: Plot of measured data and fits using the Hill formula 5.2 for enzymatic inhibition of ibuprofen and sila-ibuprofen on COX-I and COX-II based on the measurements. Error bars are the standard deviation of the measurement values.

Here Y is the inhibition of the enzyme in %, D is the upper limit for the fit, A is the lower limit, C denotes the current concentration of the test substance, IC_{50} is the concentration at which half of the inhibition is reached and n denotes the Hill coefficient, which determines the slope of the fitting curve. The fits show a similar range of inhibitory activity increase for both substances. The fitted IC_{50} for COX-I is 34 μmol L^{-1} for ibuprofen and 26 μmol L^{-1} for sila-ibuprofen. For COX-II the values are 3.3 μmol L^{-1} and 8.3 μmol L^{-1}, respectively. In general both substances were better inhibitors for COX-I than for COX-II. This is somewhat surprising since the literature reports Ibuprofen as more selective for COX-II. [479–481] However both substances have IC_{50} values in similar ranges for the two enzymes and the same selectivity, which means that sila-ibuprofen is indeed at least as potent as an inhibitor for both enzymes as Ibuprofen, while having considerably lower melting point and higher solubility.

5.2.13 Toxicological Profile

The final step before *in-vivo* investigations is the determination of toxicity of sila-ibuprofen, which was carried out by a collaboration with Prof. Dringens group, namely Erik Ehrke and Patrick Watermann. They tested different concentrations of ibuprofen and sila-ibuprofen for anti-proliferative properties and tested cell vitality using multiple models to assess the toxicity of both substances and concluded that sila-ibuprofen has similar toxicity and anti-proliferative properties like ibuprofen and therefore should in physiological concentrations not be toxic. Since this is mostly described in the publication as shown in section 5.2.14 and not work carried out for this thesis please see section 5.2.14 and appendix D.2 for details.

5.2.14 *Publication:* Sila-Ibuprofen

This publication is a big collaboration between Prof. Ralf Dringen's, Prof. Jens Beckmann's and Prof. Grabowsky's groups. I was in charge of managing the project and coordinated the efforts of all collaborators. I repeated the synthesis of sila-ibuprofen and repeated the chemical characterization. The measurement of melting enthalpy and melting point was performed by Prof. Dr. Anne Staubitz. I measured the solubility in water and performed the high resolution X-ray experiments at SPring-8 in Hyogo, Japan under supervision of beamline scientist Dr. Kunihisa Sugimoto together with Dr. Malte Fugel. The synthesis of sila-ibuprofen was repeated by Daniel Duvinage and Dr. Pim Puylaert and the resulting product given to Patrick Watermann and Eric Ehrke from Prof. Dringen's group. They performed the toxicological investigations. I performed all Quantum Crystallographic and Quantum Mechanical calculations, wrote code to perform the aNCI analysis on GPUs and the aIE investigations on a computer cluster, created the movie for the supplementary information and made all Figures and Tables in this manuscript, which was written by Prof. Beckmann, Prof. Grabowsky and me. Reprinted with permission from *Journal of Medicinal Chemistry* **2020** *63*, 21, 12614-12622. Copyright © 2020 American Chemical Society.

Journal of
Medicinal
Chemistry

pubs.acs.org/jmc

Article

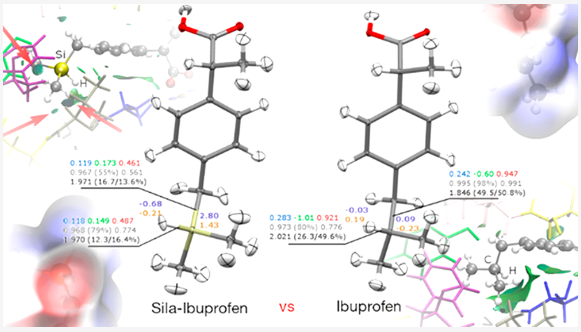
Sila-Ibuprofen

Florian Kleemiss, Aileen Justies, Daniel Duvinage, Patrick Watermann, Eric Ehrke, Kunihisa Sugimoto, Malte Fugel, Lorraine A. Malaspina, Anneke Dittmer, Torsten Kleemiss, Pim Puylaert, Nelly R. King, Anne Staubitz, Thomas M. Tzschentke, Ralf Dringen, Simon Grabowsky,* and Jens Beckmann*

Cite This: *J. Med. Chem.* 2020, 63, 12614−12622

Read Online

ABSTRACT: The synthesis, characterization, biological activity, and toxicology of sila-ibuprofen, a silicon derivative of the most common nonsteroidal anti-inflammatory drug, is reported. The key improvements compared with ibuprofen are a four times higher solubility in physiological media and a lower melting enthalpy, which are attributed to the carbon−silicon switch. The improved solubility is of interest for postsurgical intravenous administration. A potential for pain relief is rationalized via inhibition experiments of cyclooxygenases I and II (COX-I and COX-II) as well as via a set of newly developed methods that combine molecular dynamics, quantum chemistry, and quantum crystallography. The binding affinity of sila-ibuprofen to COX-I and COX-II is quantified in terms of London dispersion and electrostatic interactions in the active receptor site. This study not only shows the potential of sila-ibuprofen for medicinal application but also improves our understanding of the mechanism of action of the inhibition process.



INTRODUCTION

Ibuprofen (**1**), a nonsteroidal anti-inflammatory drug, is the gold standard in pain relief medication. It is listed in the "essential medicines list" of the World Health Organization.[1] The mechanism of action involves the inhibition of cyclooxygenase-II (COX-II), thus blocking the synthesis of prostaglandins from arachidonic acid.[2] In contrast with COX-I, which is permanently present in the body, COX-II is produced only when actual damage to the tissue or inflammation occurs. It is expressed in macrophages and synthesizes prostaglandins responsible for initial inflammation and body temperature increase. It is also expressed by endothelial cells of proliferating blood vessels and, in the case of inflammation, by endothelial cells of the hypothalamus.[3]

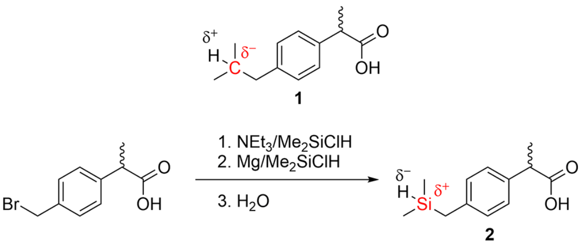
In this work, we show that ibuprofen, the preferred medication in nonsteroidal antirheumatic treatments, can still be fine-tuned and improved by the formal exchange of a carbon against a silicon atom (carbon−silicon switch).[4] We describe the synthesis of sila-ibuprofen (**2**), its full characterization, as well as toxicological and *in vitro* investigations of its pharmacological potency. The observed properties are explained using a newly developed approach based on method development in molecular dynamics (MD), quantum crystallography, and the quantum-chemical characterization of noncovalent interactions, which allows us to quantify differences and similarities between **1** and **2**.

RESULTS AND DISCUSSION

The synthesis of sila-ibuprofen (**2**) was achieved in a one-pot reaction starting from commercially available 2-[(4-bromomethyl) phenyl] propionic acid and dimethylchlorosilane, Me_2SiClH (Scheme 1).

Me_2SiClH fulfills two functions. In combination with triethylamine, NEt_3, it introduces a silyl ester that protects

Scheme 1. Lewis Formulae of Ibuprofen (1) and Sila-Ibuprofen (2) and Synthesis of 2



Received: May 13, 2020
Published: September 15, 2020

https://dx.doi.org/10.1021/acs.jmedchem.0c00813
J. Med. Chem. 2020, 63, 12614−12622

the carboxylic acid group prior to a Barbier reaction of the bromomethyl group with magnesium and Me_2SiClH. A subsequent aqueous workup deprotects the carboxylic acid and affords **2** after purification by column chromatography as a microcrystalline colorless solid in 85% yield. The ^1H and $^{13}C\{^1H\}$ NMR spectra (acetone-d_6) of **2** show the expected number of signals and confirm the purity. The $^{29}Si\{^1H\}$ NMR spectrum (acetone-d_6) of **2** shows a signal at $\delta = -11.5$ ppm with an indicative $^1J(Si{-}H)$ coupling constant of 185 Hz. The IR spectrum exhibits a characteristic signal at $\tilde{\nu} = 2132\ cm^{-1}$, which was assigned to a Si−H stretching vibration. Sila-ibuprofen (**2**) was obtained as a racemic mixture, which was used without optical resolution in this study. Although it is known that only the S-enantiomer of ibuprofen is biologically active, racemic mixtures are administered in medicinal treatments because an isomerase converts the enantiomers *in vivo*.[3]

An important issue involving the application of ibuprofen is the low solubility in physiological media, which limits the use in postsurgical intravenous treatment and is related to the rather high melting point and high melting enthalpy (Table 1).

Table 1. Key Properties of Ibuprofen (1) and Sila-Ibuprofen (2)

property	ibuprofen (1)	sila-ibuprofen (2)
melting point/°C	74−77[5]	45−47.5
melting enthalpy/kJ mol^{-1}	26.7[6]	15.8 ± 0.5
solubility (water)/mg L^{-1}	21	83 ± 3
IC_{50} (COX-I)/μM	34	26
IC_{50} (COX-II)/μM	3.3	8.3

The melting point of sila-ibuprofen (**2**, 45−47.5 °C) is significantly lower than that of ibuprofen (**1**, 74−77 °C) and even lower than that of enantiomerically pure ibuprofen (54 °C).[5] The melting enthalpy of **2** is 15.8 ± 0.5 kJ mol^{-1}, which is lower than that of **1** (26.7 kJ mol^{-1}).[6] The solubility of **2** is 83 ± 3 mg L^{-1} in water, which is approximately four times higher than that of **1**, which is 21 mg L^{-1} (Figure S1).[7] An NMR investigation of **2** under physiological conditions (pH 8, 40 °C) suggests stability over a timespan of several weeks (Figures S5 and S6).

Measurements of the IC_{50} values of inhibition of human COX-I and II by **1** or **2** in buffered saline reveal the concentrations at which 50% of the enzymatic activity is inhibited (Table 1, Figure S29). Hence, both ibuprofen (**1**) and sila-ibuprofen (**2**) more selectively inhibit COX-II than COX-I, whereas the absolute values between **1** and **2** are similar for both enzymes. A detailed discussion of how different assay conditions and parameters may affect the absolute inhibition constants and the relative COX selectivity for nonsteroidal anti-inflammatory drugs is beyond the scope of the present paper and has been discussed in detail elsewhere.[8] Hence, within the expected uncertainty of such measurements, the silicon−carbon switch has affected the inhibition properties only slightly. In summary, all properties of sila-ibuprofen (**2**) suggest improved application in physiological media while retaining a similar level of potency and mode of action compared to **1** (Table 1).

The most important metabolite of ibuprofen (**1**) is hydroxyibuprofen, which is formed upon enzymatic oxidation of the isobutyl ($CH_2Me_2\boldsymbol{C}{-}H$) group.[3] Whether the dimethylsilylmethyl ($CH_2Me_2\boldsymbol{Si}{-}H$) group will be oxidized similarly in an enzymatic reaction is an interesting scientific question. Recent

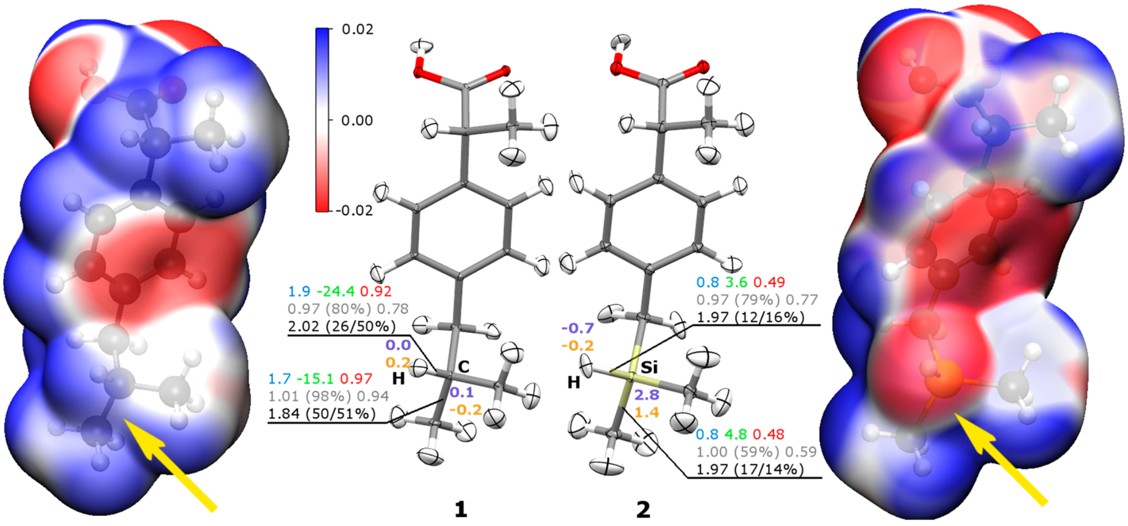


Figure 1. Electrostatic potential (in e $Å^{-1}$) of ibuprofen (**1**, left) and sila-ibuprofen (**2**, right), color-mapped onto the 0.001 au electron-density isosurface derived by XWR. The refined molecular structures are shown with anisotropic displacement parameters for all atoms, including hydrogen atoms, at a 50% probability level. XWR-derived bond descriptors and atomic charges of the regions of the molecules most affected by the umpolung are given using the following color code (references to the methods used are given in the Supporting Information): Quantum Theory of Atoms in Molecules (QTAIM) charge in e (purple), natural population analysis charge in e (orange), electron density in e $Å^{-3}$(blue) and its Laplacian in e $Å^{-5}$ (green) at the QTAIM bond critical points, delocalization index (red), natural bond orbital bond order (gray) with the percentage of covalent resonance structure derived from natural resonance theory analysis and natural localized molecular orbital NLMO/NPA bond order, and population of electron localizability indicator bond basin in e (black) with contribution from C/Si in terms of the QTAIM atomic basin volume/electron density.

https://dx.doi.org/10.1021/acs.jmedchem.0c00813
J. Med. Chem. 2020, 63, 12614−12622

residue	Ibuprofen (***1***) $\langle E_{disp/ele}\rangle$ (kJ mol^{-1})	Sila-ibuprofen (**2**) $\langle E_{disp/ele}\rangle$ (kJ mol^{-1})	wRMSD $\langle E_{Sila}\rangle$-$\langle E_{Ibu}\rangle$(disp/ele)
Arg	−17(1)/ −490(24)	−16(1)/ −490(30)	0.60/0.01
GlyAla	−41(6)/−28(4)	−35(5)/−22(4)	0.30/0.41
MetVal	−15(3)/2(4)	−21(4)/2(3)	0.49/0.07
Phe	−9(2)/1(1)	−14(4)/ −3(2)	0.49/0.76
SerLeu	−18(4)/ −20(2)	−22(5)/−27(4)	0.30/0.52
Tyr	−13(3)/ −9(5)	−11(3)/−19(7)	0.33/0.39
Σ	−113/−541	−119/−566	Δ=6/25

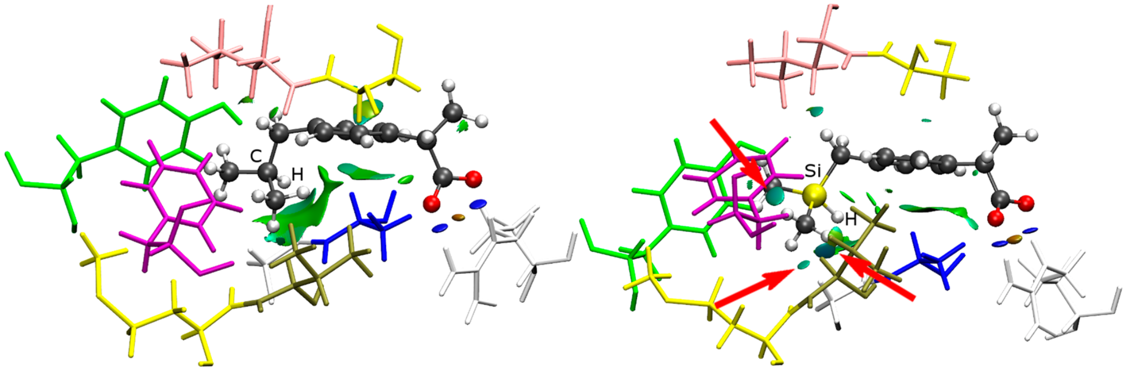


Figure 2. Visualization of residues important for close interactions inside the active site of COX-II after MD of ibuprofen (**1**, left) and sila-ibuprofen (**2**, right) (color code in the first column of the table). Visualization of the aNCI, color code: green, weak dispersion interactions; blue, stronger electrostatic interactions; orange, repulsive interactions. The table lists the corresponding pairwise interaction energies between the (sila)ibuprofen molecule and the amino acid residues, separated into dispersion and electrostatic terms,[20] but here averaged over the entire trajectory of the MD simulation. Sample standard deviations are given in brackets, and weighted root-mean-square differences are provided in a separate column. Refer to the Supporting Information for a full table of energies and a movie of the intermolecular interactions in the enzyme pocket.

work by Arnold et al. suggests that enzymes are capable of activating the artificial Si−H bond and transforming it into Si−C bonds.[9] For sila-ibuprofen (**2**), this question remains unanswered for now, but the *in situ* chemical synthesis of the potential metabolite hydroxy-sila-ibuprofen was developed and is described in the Experimental and Computational Section. Like many other silanols, hydroxy-sila-ibuprofen undergoes condensation to the related siloxane at higher concentrations.

The most significant changes relevant for the biochemical recognition of a molecule introduced by a carbon−silicon switch include the bond lengths, the molecular volume, the flexibility of the functional groups, and, most importantly, the polarization of bonds because the umpolung principle is utilized;[10] see the partial charges in Figure 1. Therefore, high-quality and high-resolution synchrotron X-ray diffraction experiments were performed on crystals of racemic mixtures of **1** and **2** (Table S1). Both compounds are isomorphous and crystallize in the monoclinic space group $P2_1/c$. Hence, molecular and solid-state structures are sterically nearly identical; the molecular volumes, determined within the electron-density isosurfaces shown in Figure 1, vary by only 7%. X-ray wavefunction refinement (XWR) was chosen as the crystallographic refinement method because it results in reliable hydrogen-atom positions based on X-ray diffraction data.[11] The tertiary H−**C**/**Si** bond length difference is 0.374 Å, whereas H−**C**/**Si**−C bond angles change by only 0.68°, averaged over three neighboring bonds.

Accurate hydrogen-atom parameters are crucial for the derivation of reliable electrostatic properties and intermolecular interaction energies.[12] It was shown before that as a first approximation, the polarization through intermolecular interactions of the biologically active compound in its crystal structure is similar to that in the enzyme.[13] The significant effect of the umpolung on the charges of the hydrogen atom by the carbon−silicon switch (q(C−***H***) = 0.0 to 0.2 *e*; q(Si−***H***) = −0.2 to −0.7 *e*, Figure 1) is reflected by the electrostatic potential (ESP) of both molecules, calculated from the experimentally constrained wave functions. **1** shows a positive ESP near the tertiary carbon atom of the isobutyl group, whereas the same area is negative around the silane hydrogen atom in **2** (Figure 1). The Politzer parameters[14] for these surfaces show a higher internal charge separation for **2** (average deviation from the mean surface potential Π = 0.0263 e Å^{-1}) compared with **1** (Π = 0.0240 e Å^{-1}); see Table S2.

Moreover, XWR models the chemical features of bonding and interactions from the experimental X-ray structure factors. Hence an experimental complementary bonding analysis is feasible.[15] Those regions that are not directly affected by the silicon−carbon switch, for example, the carboxylic acid group, show very similar intramolecular bonding features and atomic charges, but the umpolung of the Si−H bond in comparison with the C−H bond is clearly reflected by the descriptors in the direct vicinity (Figure 1, Data S3). The average electron density of the C/Si atoms and their immediate environment is a parameter complementary to the ESP and the bonding analysis. Whereas the latter reflect the polarization and governs the physical properties, the average electron density should be similar for a bioisosteric replacement.[16] Here the values are indeed similar (0.047 au (ibuprofen **1**) and 0.051 au (sila-ibuprofen **2**)) for the C/Si atom plus the directly bonded methyl/methylene groups but not as similar as for the bioisosteric tetrazole/carboxylate pair in ref 16.

In an effort to understand the effect introduced by the differences in atomic and bonding properties for the molecular recognition and the mode of action, MD simulations and subsequent quantum-chemical results averaged over the entire MD trajectory were analyzed. For this purpose, we developed new methodology and related software. (See the Experimental and Computational Section.) The available crystal structure of the complex of murine COX-II and **1** was taken from the protein crystallographic database (PDB code 4PH9)[17] and equilibrated in a solvation box with physiological sodium chloride solution. An identical procedure was applied upon the substitution of **1** by **2**. Missing force-field parameters involving the silicon atom were derived in the course of this work partially based on the XWR wave functions (Figures S12−S14, Tables S3−S5). Ibuprofen (**1**) and sila-ibuprofen (**2**) bonded to the active site of COX-II are shown in Figure 2, with the closest residues of the protein explicitly visualized. Green and blue surfaces show the noncovalent interaction (NCI) index[18] averaged over 1000 different geometries throughout the production phase of the MD (averaged noncovalent interaction index (aNCI)).[19] The type and strength of all thermally stable contacts are shown (Figure 2). In addition, thermally averaged pairwise intermolecular interaction energies are given in Figure 2.

The carboxylic acid group is the key motif for the recognition of **1** in COX-II,[3] which is reflected in Figure 2 by a large electrostatic interaction energy term with the Arg residue and by localized blue NCI surfaces. We find that the same is true for **2**, so thermal stability in the enzyme pocket is guaranteed. (See also the distance plots, Figures S15 and S16.) The interactions of the phenyl ring of both **1** and **2** with the Gly−Ala residue as well as, surprisingly, the interactions involving the C−H and Si−H groups themselves are not favorably stabilizing **2** compared with **1**. Instead, the interactions of the two methyl and the methylene groups adjacent to the silicon atom are decisive for a total stabilization of **2** relative to **1** by ∼10 kJ mol^{-1}, which includes the repulsion and polarization terms shown in Table S6 (Figures S18−S21).[20] The corresponding NCI regions (light-blue discs) representing the interactions of the methyl and methylene groups of sila-ibuprofen with the Met−Val, Phe, and Ser−Leu residues are highlighted with arrows in Figure 2. A similar calculation for **1** versus **2** binding to COX-I yields a stabilization of **1** by 19 kJ mol^{-1} (Figures S17 and S22−S25 and Table S7). Both energy differences for binding to COX-I/II (10−20 kJ mol^{-1}) are small when compared with the total binding energies of around 760−770 kJ mol^{-1} (Tables S6 and S7), indicating the similar activities of **1** and **2** against both enzymes, as confirmed by the IC$_{50}$ values in Table 1.

In free energy perturbation (FEP) calculations, a difference of the Gibbs free energy of 1.46 ± 0.14/−4.7 ± 0.2 kJ mol^{-1} was obtained when morphing ibuprofen (**1**) into sila-ibuprofen (**2**) inside COX-I/COX-II (Figures S26 and S27).[21] This confirms again that **2** is bound to COX-I and -II approximately as strongly as ibuprofen **1**; however, both theoretical approaches (FEP and averaged interaction energies, previous paragraph) show a very small stabilization of sila-ibuprofen (**2**) in COX-II but a small destabilization in COX-I. In a simplified model, the inhibition of COX-II is responsible for pain relief, whereas the inhibition of COX-I is responsible for side effects. This is promising because, in conjunction with the simple synthesis and improved solubility, it renders sila-ibuprofen (**2**) a potent candidate for drug development, with the aim of obtaining a nonsteroidal anti-inflammatory antirheumatic drug similarly potent as ibuprofen itself but with improved properties for administering the drug.

Low toxicity is a prerequisite for any pharmaceutical application. To test for potential adverse effects of sila-ibuprofen (**2**) in comparison with ibuprofen (**1**) on mammalian cells, we have used C6 glioma cells, which are widely established as model systems to study the functions and properties of brain glial cells and brain glioma.[22] The cells were exposed to **1** or **2** in concentrations of up to 1000 μM for up to 3 days. The application of 300 μM of either compound did not affect the proliferation of the cells. Furthermore, the viability of the cells was not affected by a 72 h exposure to up to 300 μM of **1** or **2** (Figure S28). These data confirm the low toxicity of ibuprofen (**1**)[3,23] and also demonstrate a low toxic potential of sila-ibuprofen (**2**), as no differences in the parameters determined were observed (Figure S28). In addition, the exposure of C6 cells to 1000 μM of **1** or **2** did not lead to any obvious change in cell morphology or to any significant increase in extracellular lactate dehydrogenase activity, demonstrating that these high concentrations were not toxic to the cells either. Serum concentrations of ibuprofen in treated patients are in the low micromolar range,[24] suggesting that in situations for *in vivo* application, no toxicity is to be expected.

■ CONCLUSIONS AND OUTLOOK

Sila-ibuprofen has similar binding characteristics and a similar inhibitory profile toward COX-I and COX-II as ibuprofen, the gold standard used for pain relief, but it has a higher solubility. This means that the carbon−silicon exchange acts as a bioisosteric replacement in the case of ibuprofen but produces beneficial physical properties.

Further studies on the ability of sila-ibuprofen to act *in vivo* as an inhibitor of COX activity as well as studies on the pharmacokinetics and pharmacodynamics of sila-ibuprofen *in vivo* are now highly desired to evaluate its pharmacological potential.

■ EXPERIMENTAL AND COMPUTATIONAL SECTION

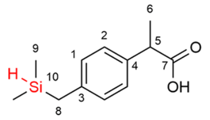
Equipment, Materials, and Methods. NMR spectra were recorded at room temperature on a Bruker Avance 600 spectrometer. ^{1}H, ^{13}C{^{1}H}, and ^{29}Si{^{1}H} NMR spectra are reported on the δ scale (ppm) and are referenced against SiMe$_4$. ^{1}H and ^{13}C{^{1}H} chemical shifts are reported relative to the residual peak of the solvent ((CD)$_3$(CD$_2$H)CO 2.09 ppm for (CD$_3$)$_2$CO) in the ^{1}H NMR spectra and to the peak of the deuterated solvent ((CD$_3$)$_2$CO 30.60 ppm) in the ^{13}C{^{1}H} NMR spectra.

The ESI (HR) MS spectra were measured on a Bruker Impact II spectrometer. Acetonitrile or dichloromethane/acetonitrile solutions ($c = 1 \times 10^{-5}$ mol L^{-1}) were injected directly into the spectrometer at a flow rate of 3 μL min^{-1}. Nitrogen was used both as a drying gas and for nebulization with flow rates of ∼5 L min^{-1} and a pressure of 5 psi. Pressure in the mass analyzer region was usually ∼1 × 10^{-5} mbar. Spectra were collected for 1 min and averaged. The nozzle-skimmer voltage was adjusted individually for each measurement.

Electronic impact mass spectroscopy (EI MS) spectra were measured on a MAT 711 spectrometer, Varian MAT. The electron energy for EI was set to 70 eV. The microanalysis was obtained from a Vario EL elemental analyzer. IR spectra were recorded with a Thermo Scientific Nicolet iS10 instrument.

Synthesis of 2-[(4-Dimethylsilylmethyl)phenyl]propionic Acid (2).[25] In a Schlenk flask under an argon atmosphere, 2-[(4-bromomethyl)phenyl]propionic acid (1.00 g, 4.11 mmol) was dissolved in diethyl ether (25 mL) and stirred at room temperature.

Dimethylchlorosilane (1.56 g, 16.4 mmol) was added, and triethylamine (0.832 g, 8.22 mmol) was added dropwise over the course of 5 min under the formation of a cloudy white solid. The reaction mixture was stirred for 24 h, and the solid was filtered and washed with dry diethyl ether (30 mL). A three-necked round-bottomed flask was equipped with a dropping funnel and a reflux condenser and charged with (0.200 g, 8.22 mmol) magnesium turnings. Under an argon atmosphere, the remaining solution was added dropwise to the magnesium turnings to obtain a constant reflux. Afterward, the suspension was refluxed for an additional 10 h. Under rapid stirring, ice water (50 mL) was added, and the organic phase was separated and worked up aqueously with distilled water (3 × 25 mL). The solvent was removed under reduced pressure to yield a sticky oil, which was purified by column chromatography by first flushing with three column volumes of *n*-hexane and subsequently eluting the product with ethyl acetate to give **2** as a microcrystalline colorless solid (0.777 g, 85% yield) after the removal of the solvent.



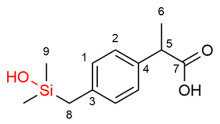
2

^{1}H NMR (acetone-d_6, 600 MHz) δ = 7.20 (d, $^{3}J(^{1}H{-}^{1}H)$ = 8.1 Hz, 2H; H-1), 7.03 (d, $^{3}J(^{1}H{-}^{1}H)$= 8.1 Hz, 2H; H-2), 3.91−3.95 (sept, $^{1}J(^{1}H{-}^{29}Si)$ = 185 Hz, $^{3}J(^{1}H{-}^{1}H)$ = 3.5 Hz, 1H; H-10), 3.68 (quart, $^{3}J(^{1}H{-}^{1}H)$ = 7.2 Hz, 1H; H-5), 2.16 (d, $^{3}J(^{1}H{-}^{1}H)$ = 3.5 Hz, 2 H; H-8), 1.40 (d, $^{3}J(^{1}H{-}^{1}H)$ = 7.2 Hz, 3 H; H-6), 0.05 (d, $^{3}J(^{1}H{-}^{1}H)$ = 3.7 Hz, 6 H; H-9). ^{13}C{^{1}H} NMR (acetone-d_6, 151 MHz) δ = 175.8 (C-7), 139.4 (C_{quart}), 138.0(C_{quart}), 129.0 ($C_{arom,}$C-1), 128.2 ($C_{arom,}$ C-2), 45.2 (C-5), 23.9 (C-8), 19.1 (C-6), −4.7 (C-9). ^{29}Si{^{1}H} NMR (acetone-d_6, 119 MHz) δ = −11.5. IR λ(Si−H): 2132 cm^{-1} (KBr-pallet). UV λ_{max}(CH_2Cl_2): 234 nm. Microanalysis calc. for $C_{12}H_{18}O_2Si$ (222.36) C, 64.82; H, 8.16; found C, 64.54; H, 8.55. EI-MS (70 eV) (*m*/*z*): 222.4 $[M]^{+}$, calculated ($C_{12}H_{18}O_2Si$) = 222.1 g/mol. HR-ESI-MS (*m*/*z*): $[M - H]^{+}$ calculated for $C_{12}H_{17}O_2Si$, 221.09923; found: 221.09926. $[M - H]^{-}$ calculated for $C_{12}H_{17}O_2Si$, 221.10033; found: 221.10017.

Metabolites of 1 and 2. Previous investigation into the metabolism of **1** exposed that it has two main metabolites that have been isolated in the urine of human subjects.[3] The characterization of these by infrared and nuclear magnetic resonance spectroscopies revealed that one of them is hydroxy-ibuprofen and the other one is the corresponding carboxylic acid.[3]

Similarly, we assume that the oxidation of **2** in the liver should produce its respective silanol derivate.[9] In this context, we have synthesized hydroxy-sila-ibuprofen (**3**) through the oxidation of compound **2** in the presence of a palladium catalyst using water as an oxidation agent. Hydroxy-sila-ibuprofen (**3**) was prepared under acidic conditions, besides its condensation product, the disiloxane **4**. Both compounds are in an equilibrium with each other.

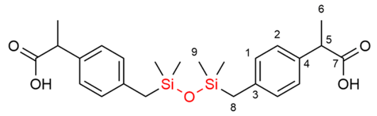
Synthesis and Characterization of 2-(4-(Hydroxydimethylsilyl-methyl)phenyl) Propionic Acid (3). **2** (500 mg, 2.25 mmol) dissolved in acetone (10.0 mL) was added to an ice-cooled suspension of Pearlman's catalyst, $Pd(OH)_2/C$ (12.0 mg), in acetone (10.0 mL) and water (0.10 mL). After the evolution of hydrogen ceased, the reaction mixture was stirred at room temperature for 10 min. The reaction mixture was filtered to remove the catalyst, and the solvent was removed from the remaining filtrate at 30 °C under reduced pressure. The removal of volatiles under reduced pressure afforded 480 mg (2.01 mmol, 89% yield) of the silanol **3** as a colorless oil.



3

^{1}H NMR (acetone-d_6, 600 MHz) δ = 7.16 (d, $^{3}J(^{1}H{-}^{1}H) = 8.2$ *Hz*, 2H; H-1), 7.05 (d, $^{3}J(^{1}H{-}^{1}H) = 8.2$ *Hz*, 2H; H-2), 3.68 (quart, $^{3}J(^{1}H{-}^{1}H) = 7.1$ *Hz*, 1H; H-5), 2.11 (s, 2H; H-8), 1.41 (d, $^{3}J(^{1}H{-}^{1}H) = 7.1$ *Hz*, 2H; H 6), 0.04 (s, 6H; H-9). ^{13}C{^{1}H} NMR (acetone-d_6, 151 MHz) δ = 175.7 (C-7), 138.7 (C_{quart}), 137.0 (C_{quart}), 128.7 (C-1), 127.5 (C-2), 44.8 (C-5), 27.8 (C-8), 18.6 (C-6), −0.94 (C-9). ^{29}Si{^{1}H} NMR (acetone-d_6, 119 MHz) δ = 14.5. EI-MS (70 eV) (*m*/*z*): 238.0 $[M]^{+\cdot}$, calculated ($C_{12}H_{18}O_3Si$) = 238.1 g/mol.

Synthesis and Characterization of the Condensation Product of 3, the Disiloxane 4. **3** (300 mg, 1.10 mmol) in acetone (100 mL) was diluted with water (12 mL), and concentrated HCl (2.50 μL) was added. The solution was left standing for 3 weeks. After this time, the volatiles were removed under reduced pressure to obtain 470 mg (1.02 mmol, 93% yield) of the disiloxane **4** as a colorless oil.



4

^{1}H NMR (acetone-d_6, 600 MHz) δ = 7.19 (d, $^{3}J(^{1}H{-}^{1}H) = 8.2$ *Hz*, H-1), 6.99 (d, $^{3}J(^{1}H{-}^{1}H) = 8.2$ *Hz*, 2H; H-2), 3.69 (quart, $^{3}J(^{1}H{-}^{1}H) = 7.1$ *Hz*, 1H; H-5), 2.08 (s, 2H; H-8), 1.42 (d, $^{3}J(^{1}H{-}^{1}H) = 7.1$ *Hz*, 2H; H-6), −0.01 (s, 6H; H-9). ^{13}C{^{1}H} NMR (acetone-d_6, 151 MHz) δ = 175.8 (C-7), 138.2 (C_{quart}), 137.1 (C_{quart}), 128.7 (C-1), 127.4 (C-2), 44.7 (C-5), 27.8 (C-8), 18.6 (C-6), −0.45 (C-9). ^{29}Si{^{1}H} NMR (acetone-d_6, 119 MHz) δ = 4.93. ESI-MS: *m*/*z* = 481.2 $[M + Na]^{+\cdot}$, calculated ($C_{24}H_{34}O_5Si_2$ + Na) = 481.6 g/mol.

Purity of the Compounds. The purity of all compounds is >95%, as determined by HPLC and NMR spectroscopy.

In detail, ibuprofen **1** was obtained commercially with a purity of ≥98%. The purity of sila-ibuprofen **2** is >95%, as determined by NMR and HPLC. Compounds **3** and **4** were investigated as metabolites and decomposition products of **2**. They are in equilibrium with each other and were not purified, and hence their biochemical properties are not subject of this Article.

Determination of the Melting Enthalpy. The melting enthalpy was determined using differential scanning calorimetry (DSC) of a sample of **1** and **2** using a Mettler-Toledo DSC3+ instrument with 40 μL aluminum crucibles with a pin (Mettler Toledo) and a pierced lid, referenced against an empty crucible with a pierced lid. The temperature program for the samples of **1** involved heating from 25 to 125 °C at 10 K min^{-1} under a flow of N_2 at 20 mL min^{-1}. The temperature program for the samples of **2** involved heating from 0 to 70 °C at 10 K min^{-1} under a flow of N_2 at 20 mL min^{-1}. Data evaluation was performed using the Software Star-e version 15.01.

Solubility Determination. The solubility was determined in an HPLC/UV experiment using an isocratic method with a 1:1 ratio of water and acetonitrile with 0.1 mol L^{-1} acetic acid as the eluent on an RP-18 gravity column detecting the absorption at 235 nm, in agreement with the UV−vis spectrum of **2**. **1** was used as an internal standard in a concentration of 0.1 mg L^{-1}, whereas equidistant calibration was done for **2** in steps of 5 mg L^{-1} starting from 5 until 55 mg L^{-1}. The resulting calibration plot is shown in Figure S1.

Determination of the Stability of Sila-Ibuprofen (2) in Solution. To investigate whether **2** is stable over time in physiological media, a solution of NaCl (0.9%) in water was used to dissolve **2** until the solution was saturated in **2**. The solution was then kept at room temperature in an NMR tube and measured every 7 days. Additionally, an identical solution was prepared and kept at 4 °C and measured every 28 days. The resulting spectra can be seen in Figures S5 and S6. These experiments show that at room temperature, the solution slowly decomposes to the disiloxane **4** over the course of 1 month. At 4 °C, decomposition is much slower; it only starts after the first month. Furthermore, the stability of **2** under physiological basic conditions (pH 8) was tested as well. A solution of $NaHCO_3$ (10 mM) in D_2O was used to dissolve sila-ibuprofen **2**. Over the course of 1 month, no decomposition by ^{1}H NMR was visible. Above a pH of 8 (tested at pH 10 and 12), sila-ibuprofen decomposes under the release of hydrogen to form the sodium salt of hydroxy-sila-ibuprofen **3**.

To complement the NMR experiments, the stability of **2** was tested on a high-performance liquid chromatography (HPLC) column under mildly acidic conditions. Purified sila-ibuprofen **2** was added onto the column, and only a single compound peak was observed after several minutes on the column for 14 repetitions of the experiment.

Crystallographic Information. Single crystals of **1** were obtained by slow evaporation of a saturated methanol solution in an open vessel. Crystals of **2** were obtained after chromatographic purification on silica gel, followed by the removal of the solvent by low-pressure evaporation and slow resublimation of **2** in a closed vessel. Information on the synchrotron measurements and pertinent crystallographic information obtained from the refinement of the structures of **1** and **2** are shown in Table S1. Data sets were measured at SPring-8, beamline BL02B1, at 25 K using a large cylindrical image plate camera.

The first step in the performed XWR is Hirshfeld atom refinement (HAR).[11] HAR uses tailor-made aspherical atomic scattering factors from a stockholder partitioning of the calculated static electron density. Here B3LYP/def2-TZVPP was used as well as a surrounding cluster of point charges and dipoles of 8 Å radius around the central molecule to simulate the crystal field. Subsequently, X-ray constrained wavefunction fitting as the second step in XWR was performed at RHF/def2-TZVPP without cluster charges to extract as much information as possible from the experimental structure factors. The program Tonto was used for the XWR procedure. From these wave functions, a 0.001 au isosurface was calculated, and the ESP was mapped onto it, resulting in the Politzer parameters given in Table S2.

Force-Field Development and Molecular Dynamics Simulations. To understand the active mode of sila-ibuprofen (**2**) in contrast with the one of conventional ibuprofen (**1**) on an atomistic scale, it was necessary to perform MD simulations. Because the parameters for a CHARMM-type force field of **2** are unknown, they had to be derived by comparing energies obtained from ab initio calculations with energies derived by the newly constructed force field. The unknown parameters of the force field are shown in Figure S12.

This procedure was carried out on the geometry and with the wave function obtained from XWR. Interaction energies with water for charge determination were calculated at a level of theory of HF/def2-TZVP, the bonded interactions were calculated at a level of theory of B3LYP/def2-TZVPP, and the torsion potential energy surface scan was performed at a level of theory of MP2/def2-SVPP. The performance of the derived force-field parameters in comparison with ab initio calculations is visualized in Figure S13, showing the dihedral potential energy surface scan energies from reference calculations and energies calculated from the force field, both normalized to their smallest value.

Additionally, the energies of three rigid potential energy surface scans of two trimethyl silane molecules, used as a smaller model compound for the silane group in **2**, in orientations showing H−Si···H−Si, H−Si···Si−H, and Si−H···H−Si contacts, were used to iteratively modify the parameters for the nonbonded interaction of the hydrogen atom of the silane group and the silicon atom itself to resemble the observed energy profiles obtained at a level of theory of B3LYP-GD3BJ/def2-TZVPP. The plots of these profiles can be seen in Figure S14. To validate these parameters in a biological context, the active site of COX-II was taken from an equilibrated structure obtained using these parameters, and the amino acids near **2** were then scanned in a radial elongation of the distance between the amino acid and **2**. The obtained bonded parameters of the force field are shown in Table S3. The parameters for the Lennard-Jones potential are shown in Table S4, and charges assigned to the individual types of atoms in both **1** and **2** can be seen in Table S5.

The MD simulations and all derived properties (Supporting Information Sections 4−6) are based on available crystal structures of ibuprofen bonded to the active sites of COX-I and COX-II. Unfortunately, the crystal structures of ibuprofen with human COX do not exist, so we had to resort to crystal structures of complexes of ibuprofen with ovine COX-I (PDB code 1EQG)[26] and murine COX-II (PDB code 4PH9).[17] Human and animal COX are pharmacologically not identical,[27] but they are by far the best models available for our study.

The derived force-field parameters were used in addition to the parameters obtained for **1** from the SwissParam service,[28] to describe ibuprofen (**1**) and sila-ibuprofen (**2**), whereas a CHARMM force field[29] was applied for the protein, sugars, and heme residues to perform MDs for 400 ns on each complex inside a 110 × 110 × 110 Å^3 box, including explicit water molecules (TIP3P) and sodium chloride ions corresponding to a concentration of 0.15 mol L^{-1} using *NAMD2*.[30] The time step chosen for the simulations was 1 fs at a temperature of 300 K and a pressure of 1 atm used for prior equilibration. Periodic boundary conditions were used. After the equilibration, the system was simulated without a thermostat or barostat to ensure that no outer influence caused changes in the binding of the drug molecule or conformational changes of the protein. A plot of the distance of both oxygen atoms of the carboxylic acid group to the corresponding arginine hydrogen bond donors is shown in Figures S15 and S16.

Averaged Noncovalent Interaction Index. The general idea of the aNCI was introduced by Wu et al.,[19] who made their source code available. Because this approach needs many different evaluations of the NCI for the different geometries of the MD, the computational cost of this procedure is very high. For a system with a size of proteins and for long runs, this becomes too demanding to be done on a reasonable time scale. Rubez et al. wrote a kernel that performs the calculation of promolecular NCI calculations highly parallelized on graphics cards (cuNCI).[31] This source code is also available. Because during an MD no wave function is available, only promolecular calculations are possible; therefore, this kernel was optimal to extend the general functionality of the graphics cards code by the averaging procedure proposed by Wu et al.[19] The resulting program plug-in is called acuNCI in reference to both previous programs. The gradient and electron density are calculated numerically on a grid and averaged after each volumetric data set. In principle, this approach would also be possible for wave-function-based calculations and could be done using a similar kernel, using wave functions obtained by QM/MM calculations. The program will be available free of charge, also including wave-function-based calculated property files, where the calculation of numerical volumetric data is performed on graphics cards. The results of these calculations are shown in Movie S1, moving the viewer through the protein and binding pocket of COX-II in 3D.

Averaged Interaction Energies. To quantify the differences in the interactions between amino acid residues and sila-ibuprofen (**2**) in contrast with ibuprofen (**1**), the simulations for the aNCI plots were used to calculate the interaction energies using the program Tonto (commit 5ba65f7 on GitHub, https://github.com/dylan-jayatilaka/tonto), which is the backend of CrystalExplorer,[32] which is known for calculations of interaction energies and energy frameworks.[20,33] Because amino acids in proteins are part of a bigger molecule, it was necessary to saturate the bonds with hydrogen atoms that were cut when extracting individual residue coordinates from the protein. Hydrogen atoms were added using internal *z*-matrix notation for the determination of the positions using a bond length of 1.07 Å for hydrogen atoms bonded to the N-terminus and 1.00 Å for the C-terminal hydrogen atoms. sp^2 hybridization of the corresponding atom was assumed, using an ideal angle of 120° for the H−N/C−C$_\alpha$ angle and 180° for the dihedral angle using the carboxy-oxygen. A script to automatically calculate wave functions at a level of theory of B3LYP/6-31G(d,p) for the resulting amino acids was used to calculate wave functions for 1000 different frames of all 13 amino acids, each 1 ps apart, during the MD for both simulations in COX-II and for 12 residues with identical timing in COX-I. Each wave function was then analyzed using Tonto and the four terms of the interaction energy as well as the total energy are plotted in Figures S18−S21 for COX-II and Figures S22−S25 for COX-I. The averaged values as well as standard uncertainties are shown in Tables S6 and S7, respectively. The residues that had the closest distance around the silicon function and were used to define the energy difference given in the Results and Discussion section (Figure 2) are highlighted by color in these tables.

Free Energy Perturbation Calculations. Using the FEP method[21] in NAMD2[30] using ParseFEP,[34] the calculation of a so-called "alchemical" transformation is possible, which describes in this case the gradual exchange of the C atom in ibuprofen by Si and the parameters associated with this as well as the neighboring atoms, as affected by the change of the force field. In principle, both atoms are always present in the calculation, whereas their contribution to the system is weighted by the multiplicator λ, which is changed throughout the simulation from 0.0 to 1.0 in an interval size of 0.025, with 100 000 time steps for the equilibration of the system prior to the evaluation of the FEP density of states as well as energetics during the following 500 000 time steps, before increasing the λ interval once more. The resulting plots for the forward ($\lambda \in [0.0, 1.0]$) and backward ($\lambda \in [1.0, 0.0]$) transformation of the density of states and the convergence of energy in the system are plotted in Figures S26 and S27.

Cell Toxicological Investigation. To test for potential adverse effects of **1** or **2** on mammalian cells, we have used C6 glioma cells, which are widely used as a model system to study the functions and properties of brain glial cells and brain glioma.[22,35] The cells were exposed to either **1** or **2** in concentrations of up to 1000 μM for up to 3 days. The application of 300 μM of these compounds did not affect the proliferation of the cells, as demonstrated by the absence of any significant difference in the increase in cellular activity of the enzyme lactate dehydrogenase (LDH) compared with control cells (Figure S28A,B). Furthermore, the viability of the cells was not affected by a 72 h exposure to up to 300 μM of **1** or **2**, as indicated by the absence of any significant increase in the extracellular LDH activity (Figure S28C), by the, at best, low loss in cellular protein per well (Figure S28D), by the unaltered water soluble tetrazolium salt-1 (WST-1) reduction capacity (Figure S28E), and by the almost unaltered cellular lactate production (Figure S28F). These data confirm the low toxic potential of **1**[3,24] and also demonstrate a low toxic potential of **2**, as no differences in the parameters determined were observed for cells that had been treated with either **1** or **2** in concentrations of 100 or 300 μM (Figure S28A−F).

Moreover, the exposure of C6 cells to 1000 μM of **1** or **2** did not lead to any obvious change in cell morphology (data not shown) or to any significant increase in extracellular LDH, demonstrating that these high concentrations were not toxic to the cells either. However, at a concentration of 1000 μM, both compounds drastically lowered the cell proliferation (Figure S28A,B). Concerning this antiproliferative effect, 1000 μM of **2** appeared to have a slightly higher potential than 1000 μM of **1**, as indicated by the significantly lower values determined for the cellular LDH activity (Figure S28A,B), the cellular protein content (Figure S28D), as well as the cellular WST1 reduction capacity (Figure S28E). Nevertheless, it should be considered that the serum concentrations of **1** in treated patients are in the low micromolar range,[24] suggesting that the antiproliferative effect observed for very high concentrations of **1** or **2** will not be relevant for an *in vivo* situation. A potential reason for the antiproliferative action of **1** applied in high concentrations may be its reported side effect of uncoupling the mitochondrial respiratory chain,[36] which will diminish the mitochondrial ATP production and thereby slow down the cell proliferation.

Enzyme Activity Measurements to Determine IC_{50} Values. Inhibition studies of COX-I and COX-II and the determination of the IC_{50} of the enzymes by **1** and **2** were performed by the company Eurofins Cerep (Le Bois L'Êveque, France) according to ref 37 using human recombinant COX-I and COX-II enzymes from Sf9 cells in buffered saline. The test substrates applied were 1.2 μM of arachidonic acid and 25 μM of ADHP, and the incubation times were 3 min (COX-I) and 5 min (COX-II). The activity was measured by monitoring the resorfurin content as a measure for activity quantified using fluorimetry. Concentrations were selected in half-logarithmic steps in a range from 0.1 to 100 μM of test substance. The resulting % inhibitions are given in Table S8 and are visualized in Figure S29 with two reference substances for each enzyme. Regression of the data was performed using the model

$$Y = D + \left(\frac{A - D}{1 + \left(\frac{C}{IC_{50}} \right)^{n}} \right)$$

employing the commercial software SigmaPlot 4.0 and Hill software (Y = activity, A = left asymptote, D = right asymptote, C = concentration, IC_{50} = concentration of half-maximal inhibition, and n = Hill coefficient).

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jmedchem.0c00813.

Details of characterization, quantum crystallography, force-field development, the aNCI index, the averaged interaction energies, FEP calculations, toxicological investigations, enzyme activity measurements to determine IC_{50} values, references to methods, and software used in the bonding analysis (PDF)

Movie, Data 002. Guided visual representation of the sila-ibuprofen−COX-II complex with aNCI isosurfaces and a comparison with the aNCI of the ibuprofen−COX-II complex (MP4)

Data 003. Bond lengths of MD simulations (XLSX)

Data 004. Bond properties of complementary bonding analyses (XLSX)

Data 005. Interaction energies for MD simulations of COX-I and COX-II complexes (XLSX)

Data 006. Names, SMILES notation, and IC_{50} values of ibuprofen **1** and sila-ibuprofen **2** (CSV)

Data 007. CIF report of **1** (CIF)

Data 008. X-ray structure factors of **1** (CIF)

Data 009. CIF report of **2** (CIF)

Data 010. X-ray structure factors of **2** (CIF)

■ AUTHOR INFORMATION

Corresponding Authors

Jens Beckmann − *University of Bremen, Institute for Inorganic Chemistry and Crystallography, 28359 Bremen, Germany; Free University of Berlin, Institute of Chemistry and Biochemistry, 14195 Berlin, Germany;* orcid.org/0000-0002-8548-1821; Email: j.beckmann@uni-bremen.de

Simon Grabowsky − *University of Bremen, Institute for Inorganic Chemistry and Crystallography, 28359 Bremen, Germany; University of Bern, Department of Chemistry and Biochemistry, 3012 Bern, Switzerland;* orcid.org/0000-0002-3377-9474; Email: simon.grabowsky@dcb.unibe.ch

Authors

Florian Kleemiss − *University of Bremen, Institute for Inorganic Chemistry and Crystallography, 28359 Bremen, Germany; University of Bern, Department of Chemistry and Biochemistry, 3012 Bern, Switzerland;* orcid.org/0000-0002-3631-1535

Aileen Justies − *Free University of Berlin, Institute of Chemistry and Biochemistry, 14195 Berlin, Germany*

Daniel Duvinage − *University of Bremen, Institute for Inorganic Chemistry and Crystallography, 28359 Bremen, Germany;* orcid.org/0000-0002-9387-7172

Patrick Watermann − *University of Bremen, Center for Biomolecular Interactions Bremen and Center for*

Environmental Research and Sustainable Technology, 28359 Bremen, Germany

Eric Ehrke − *University of Bremen, Center for Biomolecular Interactions Bremen and Center for Environmental Research and Sustainable Technology, 28359 Bremen, Germany*

Kunihisa Sugimoto − *Japan Synchrotron Radiation Research Institute (JASRI), Diffraction & Scattering Division, Sayo-gun, Hyogo 679-5198, Japan; Institute for Integrated Cell-Material Sciences (iCeMS), Kyoto University, Kyoto 606-8501, Japan;* orcid.org/0000-0002-0103-8153

Malte Fugel − *University of Bremen, Institute for Inorganic Chemistry and Crystallography, 28359 Bremen, Germany;* orcid.org/0000-0002-5830-0298

Lorraine A. Malaspina − *University of Bremen, Institute for Inorganic Chemistry and Crystallography, 28359 Bremen, Germany; University of Bern, Department of Chemistry and Biochemistry, 3012 Bern, Switzerland;* orcid.org/0000-0002-8281-4264

Anneke Dittmer − *University of Bremen, Institute for Inorganic Chemistry and Crystallography, 28359 Bremen, Germany*

Torsten Kleemiss − *University of Bremen, Institute for Inorganic Chemistry and Crystallography, 28359 Bremen, Germany*

Pim Puylaert − *University of Bremen, Institute for Inorganic Chemistry and Crystallography, 28359 Bremen, Germany;* orcid.org/0000-0003-3397-2677

Nelly R. King − *Free University of Berlin, Institute of Chemistry and Biochemistry, 14195 Berlin, Germany*

Anne Staubitz − *University of Bremen, Institute for Analytical and Organic Chemistry, 28359 Bremen, Germany;* orcid.org/0000-0002-9040-3297

Thomas M. Tzschentke − *Independent Pharma Consultant, 52080 Aachen, Germany*

Ralf Dringen − *University of Bremen, Center for Biomolecular Interactions Bremen and Center for Environmental Research and Sustainable Technology, 28359 Bremen, Germany;* orcid.org/0000-0001-7869-1305

Complete contact information is available at:
https://pubs.acs.org/10.1021/acs.jmedchem.0c00813

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

In memory of the late Professor Werner Reutter, Charité Berlin. We thank Lucio Colombi Ciacchi and Steffen Lid for helpful discussions on the force field development. The synchrotron-radiation experiments were performed on the BL02B1 beamline of SPring-8 with the approval of the Japan Synchrotron Radiation Research Institute (JASRI) under proposal no. 2017A1233. Funding by Studienstiftung des Deutschen Volkes (Promotionsstipendium); Deutsche Forschungsgemeinschaft: GR 4451/1-1, GR 4451/2-1, and BE 3716/7-1; Forschungskommission Freie Universität Berlin; and Forschungskommission Universität Bremen (Brückenstipendium) is acknowledged.

ABBREVIATIONS

XWR, X-ray wavefunction refinement; QTAIM, quantum theory of atoms in molecules; NLMO, natural localized molecular orbital; NPA, natural population analysis; ESP, electrostatic potential; NCI, noncovalent interaction; aNCI, averaged noncovalent interaction; FEP, free energy perturbation; DSC, differential scanning calorimetry; HAR, Hirshfeld atom refinement; LDH, lactate dehydrogenase

REFERENCES

(1) *21st WHO Essential Medicines List (EML)*; World Health Organization, 2019. www.who.int/medicines/publications/essentialmedicines (accessed August 6, 2020).

(2) Prusakiewicz, J. J.; Duggan, K. C.; Rouzer, C. A.; Marnett, L. J. Differential sensitivity and mechanism of inhibition of COX-2 oxygenation of arachidonic acid and 2-arachidonoylglycerol by ibuprofen and mefenamic acid. *Biochemistry* **2009**, *48*, 7353−7355.

(3) Rainsford, K. D. *Ibuprofen: Discovery, Development and Therapeutics*; Wiley-Blackwell: Hoboken, NJ, 2015.

(4) (a) Franz, A.; Wilson, S. O. Organosilicon molecules with medicinal applications. *J. Med. Chem.* **2013**, *56*, 388−405. (b) Ramesh, R.; Reddy, D. S. Quest for novel chemical entities through incorporation of silicon in drug scaffolds. *J. Med. Chem.* **2018**, *61*, 3779−3798.

(5) Romero, A.; Rhodes, C. Stereochemical aspects of the molecular pharmaceutics of ibuprofen. *J. Pharm. Pharmacol.* **1993**, *45*, 258−262.

(6) Xu, F.; Sun, L.-X.; Tan, Z.-C.; Liang, J.-G.; Li, R.-L. Thermodynamic study of ibuprofen by adiabatic calorimetry and thermal analysis. *Thermochim. Acta* **2004**, *412*, 33−57.

(7) Yalkowsky, S. H.; He, Y.; Jain, P. *Handbook of Aqueous Solubility Data*; CRC Press: Boca Raton, FL, 2016.

(8) (a) Riendeau, D.; Charleson, S.; Cromlish, W.; Mancini, J. A.; Wong, E.; Guay, J. Comparison of the cyclooxygenase-1 inhibitory properties of nonsteroidal anti-inflammatory drugs (NSAIDs) and selective COX2-inhibitors, using sensitive microsomal and platelet assays. *Can. J. Physiol. Pharmacol.* **1997**, *75*, 1088−1095. (b) Riendeau, D.; Percival, M. D.; Brideau, C.; Charleson, S.; Dubé, D.; Ethier, D.; Falgueyret, J.-P.; Friesen, R. W.; Gordon, R.; Greig, G.; Guay, J.; Mancini, J.; Ouellet, M.; Wong, E.; Xu, L.; Boyce, S.; Visco, D.; Girard, Y.; Prasit, P.; Zamboni, R.; Rodger, I. W.; Gresser, M.; Ford-Hutchinson, A. W.; Young, R. N.; Chan, C.-C. Etoricoxib (MK-0663): Preclinical profile and comparison with other agents that selectively inhibit cyclooxygenase-2. *J. Pharmacol. Exp. Ther.* **2001**, *296*, 558−566. (c) Kato, M.; Nishida, S.; Kitasato, H.; Sakata, N.; Kawai, S. Cyclooxygenase-1 and cyclooxygenase-2 selectivity of non-steroidal anti-inflammatory drugs: investigation using human peripheral monocytes. *J. Pharm. Pharmacol.* **2001**, *53*, 1679−1685.

(9) (a) Kan, S. B. J.; Lewis, R. D.; Chen, K.; Arnold, F. H. Directed evolution of cytochrome c for carbon-silicon bond formation: Bringing silicon to life. *Science* **2016**, *354*, 1048−1051. (b) Bähr, S.; Brinkmann-Chen, S.; Garcia-Borras, M.; Roberts, J. M.; Katsoulis, D. E.; Houk, K. N.; Arnold, F. H. Selective Enzymatic Oxidation of Silanes to Silanols. *Angew. Chem., Int. Ed.* **2020**, *59*, 15507−15511.

(10) Seebach, D. Methods of reactivity umpolung. *Angew. Chem., Int. Ed. Engl.* **1979**, *18*, 239−258.

(11) (a) Woińska, M.; Grabowsky, S.; Dominiak, P. M.; Woźniak, K.; Jayatilaka, D. Hydrogen atoms can be located accurately and precisely by x-ray crystallography. *Sci. Adv.* **2016**, *2*, e1600192. (b) Woińska, M.; Jayatilaka, D.; Dittrich, B.; Flaig, R.; Luger, P.; Woźniak, K.; Dominiak, P. M.; Grabowsky, S. Validation of x-ray wavefunction refinement. *ChemPhysChem* **2017**, *18*, 3334−3351.

(12) Müller-Dethlefs, K.; Hobza, P. Noncovalent interactions: A challenge for experiment and theory. *Chem. Rev.* **2000**, *100*, 143−167.

(13) Mladenovic, M.; Arnone, M.; Fink, R. F.; Engels, B. Environmental effects on charge densities of biologically active molecules: Do molecule crystal environments indeed approximate protein surroundings? *J. Phys. Chem. B* **2009**, *113*, 5072−5082.

(14) Politzer, P.; Murray, J. S.; Peralta-Inga, Z. Molecular surface electrostatic potentials in relation to noncovalent interactions in biological systems. *Int. J. Quantum Chem.* **2001**, *85*, 676−684.

(15) Fugel, M.; Malaspina, L. A.; Pal, R.; Thomas, S. P.; Shi, M. W.; Spackman, M. A.; Sugimoto, K.; Grabowsky, S. Revisiting a historic concept using experimental quantum crystallography: Are phosphate,

sulfate and perchlorate anions hypervalent? *Chem. - Eur. J.* **2019**, *25*, 6523−6532.

(16) Matta, C. F.; Arabi, A. A.; Weaver, D. F. The bioisosteric similarity of the tetrazole and carboxylate anions: Clues from the topologies of the electrostatic potential and of the electron density. *Eur. J. Med. Chem.* **2010**, *45*, 1868−1872.

(17) Orlando, B. J.; Lucido, M. J.; Malkowski, M. G. The structure of ibuprofen bound to cyclooxygenase-2. *J. Struct. Biol.* **2015**, *189*, 62−66.

(18) Johnson, E. R.; Keinan, S.; Mori-Sanchez, P.; Contreras-Garcia, J.; Cohen, A. J.; Yang, W. Revealing noncovalent interactions. *J. Am. Chem. Soc.* **2010**, *132*, 6498−6506.

(19) Wu, P.; Chaudret, R.; Hu, X.; Yang, W. Noncovalent interaction analysis in fluctuating environments. *J. Chem. Theory Comput.* **2013**, *9*, 2226−2234.

(20) Turner, M. J.; Grabowsky, S.; Jayatilaka, D.; Spackman, M. A. Accurate and efficient model energies for exploring intermolecular interactions in molecular crystals. *J. Phys. Chem. Lett.* **2014**, *5*, 4249−4255.

(21) Dixit, S. B.; Chipot, C. Can absolute free energies of association be estimated from molecular mechanical simulations? The biotin-streptavidin system revisited. *J. Phys. Chem. A* **2001**, *105*, 9795−9799.

(22) Lenting, K.; Verhaak, R.; ter Laan, M.; Wesseling, P.; Leenders, W. Glioma: experimental models and reality. *Acta Neuropathol.* **2017**, *133*, 263−282.

(23) Berns, M.; Toennessen, M.; Koehne, P.; Altmann, R.; Obladen, M. Ibuprofen augments bilirubin toxicity in rat cortical neuronal culture. *Pediatr. Res.* **2009**, *65*, 392−396.

(24) Davies, N. M. Clinical pharmacokinetics of ibuprofen. *Clin. Pharmacokinet.* **1998**, *34*, 101−154.

(25) Beckmann, J.; Justies, A. Silicon Containing Compounds and Their Use as Anti-Inflammatory Agents. PCT Int. Appl. WO 2009004063 A1, 2009. Beckmann, J.; Justies, A. Silicon Containing Compounds and Their Use as Anti-Inflammatory Agents. Eur. Pat. Appl. EP 2011502 A1, 2009.

(26) Selinsky, B. S.; Gupta, K.; Sharkey, C. T.; Loll, P. J. Structural analysis of NSAID binding by prostaglandin H2 synthase: time-dependent and time-independent inhibitors elicit identical enzyme conformations. *Biochemistry* **2001**, *40*, 5172−5180.

(27) Ramesha, C. S. Human and Rat Cyclooxygenases are Pharmacologically Distinct. In *Eicosanoids and Other Bioactive Lipids in Cancer, Inflammation, and Radiation Injury 3*; Springer: Boston, MA, 1997; pp 67−71.

(28) Zoete, V.; Cuendet, M. A.; Grosdidier, A.; Michielin, O. SwissParam: a fast force field generation tool for small organic molecules. *J. Comput. Chem.* **2011**, *32*, 2359−2368.

(29) (a) MacKerell, A. D.; Feig, M.; Brooks, C. L. Extending the treatment of backbone energetics in protein force fields: Limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations. *J. Comput. Chem.* **2004**, *25*, 1400−1415. (b) MacKerell, A. D.; Bashford, D.; Bellott, M.; Dunbrack, R. L.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E.; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiorkiewicz-Kuczera, J.; Yin, D.; Karplus, M. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B* **1998**, *102*, 3586−3616.

(30) (a) Kalé, L.; Skeel, R.; Bhandarkar, M.; Brunner, R.; Gursoy, A.; Krawetz, N.; Phillips, J.; Shinozaki, A.; Varadarajan, K.; Schulten, K. NAMD2: greater scalability for parallel molecular dynamics. *J. Comput. Phys.* **1999**, *151*, 283−312. (b) Phillips, J.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kalé, L.; Schulten, K. Scalable molecular dynamics with NAMD. *J. Comput. Chem.* **2005**, *26*, 1781−1802.

(31) Rubez, G.; Etancelin, J. M.; Vigouroux, X.; Krajecki, M.; Boisson, J. C.; Hénon, E. GPU accelerated implementation of NCI calculations using promolecular density. *J. Comput. Chem.* **2017**, *38*, 1071−1083.

(32) Turner, M. J.; McKinnon, J. J.; Wolff, S. K.; Grimwood, D. J.; Spackman, P. R.; Jayatilaka, D.; Spackman, M. A. *Crystal Explorer 17*; University of Western Australia, 2017. http://hirshfeldsurface.net.

(33) (a) Mackenzie, C. F.; Spackman, P. R.; Jayatilaka, D.; Spackman, M. A. CrystalExplorer model energies and energy frameworks: extension to metal coordination compounds, organic salts, solvates and open-shell systems. *IUCrJ* **2017**, *4*, 575−587. (b) Turner, M. J.; Thomas, S. P.; Shi, M. W.; Jayatilaka, D.; Spackman, M. A. Energy frameworks: insights into interaction anisotropy and the mechanical properties of molecular crystals. *Chem. Commun.* **2015**, *51*, 3735−3738.

(34) Liu, P.; Dehez, F.; Cai, W.; Chipot, C. A. A toolkit for the analysis of free-energy perturbation calculations. *J. Chem. Theory Comput.* **2012**, *8*, 2606−2616.

(35) (a) Barth, R. F.; Kaur, B. Rat brain tumor models in experimental neuro-oncology: the C6, 9L, T9, RG2, F98, BT4C, RT-2 and CNS-1 gliomas. *J. Neuro-Oncol.* **2009**, *94*, 299−312. (b) Grobben, B.; De Deyn, P. P.; Slegers, H. Rat C6 glioma as experimental model system for the study of glioblastoma growth and invasion. *Cell Tissue Res.* **2002**, *310*, 257−270.

(36) Browne, G. S.; Nelson, C.; Nguyen, T.; Ellis, B. A.; Day, R. O.; Williams, K. M. Stereoselective and substrate-dependent inhibition of hepatic mitochondrial *β*- oxidation and oxidative phosphorylation by the non-steroidal anti-inflammatory drugs ibuprofen, flurbiprofen, and ketorolac. *Biochem. Pharmacol.* **1999**, *57*, 837−844.

(37) Vanachayangkul, P.; Tolleson, W. H. Inhibition of heme peroxidases by melamine. *Enzyme Res.* **2012**, *2012*, 1−7.

5.3 Crystal and Enzyme Environmental Effect on Ibuprofen and Sila-Ibuprofen

The study presented in section 5.2 showed the lower melting enthalpy and better solubility of the biosiostere silicon-carbon exchanged equivalent of ibuprofen, sila-ibuprofen. Given this similarity in the *in-vitro* tests the similarity of both crystal structures of ibuprofen and sila-ibuprofen in comparison to their situation in the enzyme simulations is of interest. Therefore a similar methodology to analyse the environmental effects on the drug molecule as performed in section 5.1 might yield insights into the effect the *umpolung* has in terms of environmental impacts on the density and binding situation of both molecules and whether the observed changes in the dipole moment of the silane functional group have a bigger impact on the responsiveness to environmental influences of sila-ibuprofen compared to ibuprofen.

In biological media, only the S-enantiomer of ibuprofen is active. Even if the R-enantiomer can be converted into the S-enantiomer quite readily, the docking in COX enzymes only occurs with the S-enantiomer. [299] Also, the actively binding molecule inside COX enzymes is the deprotonated anionic form of ibuprofen, therefore a direct comparison between the crystal structure of pure ibuprofen and sila-ibuprofen and their corresponding situation in the enzyme is not possible.

For the direct comparison of the environmental effects on both molecules to the situation in the enzyme the separation of enantiomers is required. Also, the crystal structure to be used for the comparison of the environmental effects should contain the deprotonated anion. The following sections will explain how an enantiomeric separation was achieved and how the analysis of the environmental effects on both ibuprofen and sila-ibuprofen was conducted.

5.3.1 Separation of Enantiomers & Co-Crystalization with Arginine

The enantiomeric separation of both ibuprofen and sila-ibuprofen was attempted using a procedure reported in the literature, where either R- or S-1-phenylethyl-1-amine (PEA) is added to a solution of ibuprofen to form only co-crystals of one of the enantiomers. [482] The crystal structure of the co-crystal with ibuprofen is reported in the literature. [483, 484] For sila-ibuprofen, only crystals of very low quality could be obtained. Both crystals (ibuprofen and sila-ibuprofen co-crystallized with PEA) were measured at SPring-8 at beamline BL02-B1 by Malte Fugel and myself to yield datasets for further analysis. The structures could be solved and refined using HAR in NoSpherA2; anomalous dispersion values for the employed wavelength were used from the Sasaki tables. [485] The final geometries are visualized in Figure 5.35 and the refinement statistics are given in Table 5.8. A level of theory of PBE/def2-TZVPP was used in ORCA with normal integration accuracy and normal SCF settings.

It was found that the enantiomeric separation was not complete for the ibuprofen-PEA compound. A small amount of the other enantiomer was found in the crystal structure, which could be modelled by a disorder refinement. The neighbouring ordered atoms were split but treated using EXYZ and EADP constraints (compare section 3.1.2), so that two types of scattering factors could be calculated, accounting for the different bonding situations these atoms are in. In the ibuprofen-PEA co-crystal, the occupancy of the S-enantiomer was refined to be 0.188(7), which means there is 81.2(7)% R-enantiomer in the crystal structure.

The resolution of the measured dataset of the co-crystal of sila-ibuprofen and PEA is much lower and not all hydrogen atom positions could be refined freely (compare also Table 5.8). The hydrogen atoms of the CH_2 group of sila-ibuprofen were given identical U_{iso} values in both parts, the methyl groups attached to the silicon atom were also refined using a free variable to define their U_{iso} values. All C-H distances were fixed using the corresponding AFIX commands and using reported neutron diffraction averaged distances for the functional group. [486] The N-H distance was refined freely and the Si-H distance of the silane functional group was fixed to the distance obtained from the crystal structure of sila-ibuprofen (see section 5.2). The disorder in the dimethyl-silane functional group was treated by creating two parts that were rotated to match the residual density peaks. A disorder refinement of the enantiomeric purity of the sila-ibuprofen was attempted – as has been done for ibuprofen-PEA – but unfortunately the refinement was unstable due to the low quality of the available data and therefore only one enantiomer, the S-enantiomer, was used during the refinement. The residual density level is comparatively low and therefore no real judgement on the presence of the other enantiomer was possible with the given data quality.

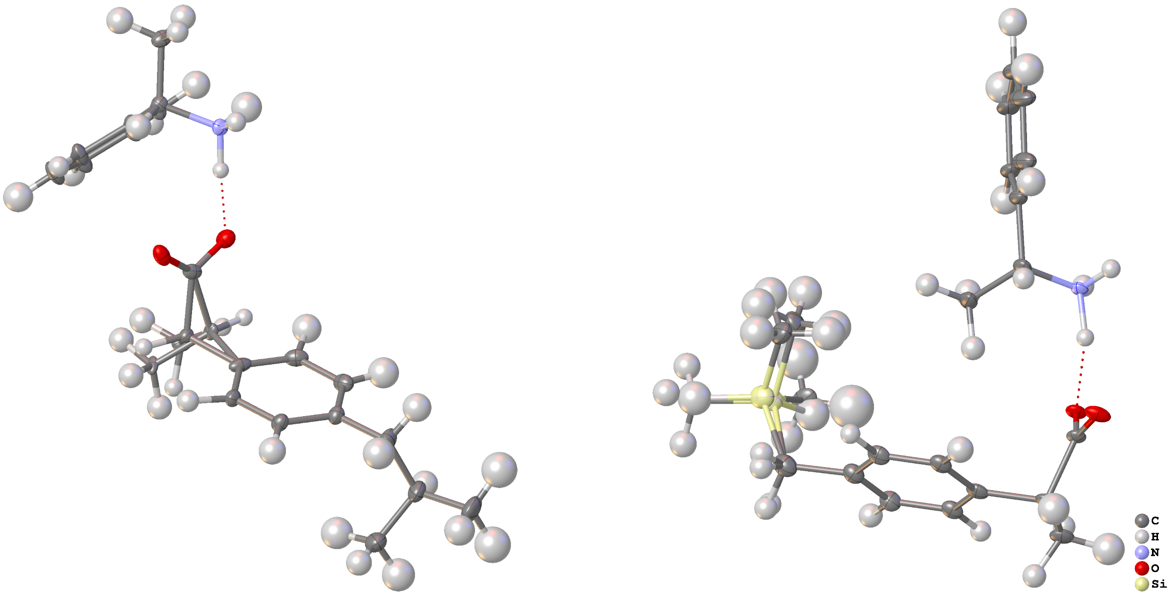


Figure 5.35: Visualization of the HAR structures of the ibuprofen (left) and sila-ibuprofen (right) co-crystals formed with PEA. Atomic displacement parameters are drawn at 50% probability level. Both disorder parts are shown for both structures in the Figure. In the ibuprofen structure the R-enantiomer has 81.2(7)% occupancy. In sila-ibuprofen, the disorder of the methyl groups has 47.9 (11)% occupancy of PART 1 and only the S-enantiomer is found. The co-crystals of ibuprofen were formed with S-PEA, while sila-ibuprofen co-crystals were formed with R-PEA.

A HAR of these datasets would never have been possible with the HAR implementation in *Tonto*. Neither disorder, restraints, riding models nor the use of adjustable (e.g. *shelx*-type) weighting schemes are possible there, so these refinements put to practical use most of the possibilities that *NoSpherA2* has to offer. Advanced structural refinements like HAR can only be done on these structures using the new implementation of HAR in *NoSpherA2* developed and presented during this work.

Table 5.8: Crystallographic measurement and refinement statistics of ibuprofen and sila-ibuprofen PEA co-crystals after HAR using *NoSpherA2*.

Structure	Ibuprofen-PEA	Sila-ibuprofen-PEA
Space group	$P\,2_12_12_1$	
a /Å	5.9130(12)	6.8160(14)
b /Å	15.305(3)	12.721(3)
c /Å	22.257(4)	23.613(5)
V /Å^3	2014.2(7)	2047.4(7)
T /K	25	
d_{min} /Å	0.70	0.80
$\lambda_{\text{X-ray}}$ /Å	0.3567	
R_{int}	0.0552	0.0517
Avg. redundancy	3.82	4.15
Completeness	1.00	1.00
Average I/σ	29.8	14.3
# of refln. measured	87404	17320
# of unique refln.	6129	4173
Obs. Criterium	$I_o >= 2\sigma(I_o)$	
# of observed refln.	5977	2694
Weighting scheme w=	$\frac{1}{\sigma^2(F_o)+(0.0848P)^2+2.4386P}^*$	$\frac{1}{\sigma^2(F_o)+(0.0739P)^2+1.1599P}^*$
# of Parameters	365	261
# of Restraints/Constraints	18/6	0/27
N_p/N_{ref}	16.8	15.9
Final R_1	0.0722	0.0802
Final $R_{1,\text{all}}$	0.0736	0.1176
Final wR_2	0.1849	0.1854
Flack [487]$^+$	-1(1)	-0.1(4)
Max $\Delta\rho$ /eÅ^{-3}	0.497	0.338
Min $\Delta\rho$ /eÅ^{-3}	-0.314	-0.204
CSD depostion number		

$^*\ P = \frac{F_o^2+2F_c^2}{3}$; $^+$ There is an error in the calculation of the Flack parameter [487] using *olex2.refine*, therefore parameters obtained using *shelx* are reported.

The PEA-ibuprofen and PEA-sila-ibuprofen could be understood as a model of the interaction of the small drug molecules with an amine function, similar to the binding with the guanidine functional group of arginine inside COX-I and COX-II. To produce a more direct model of the interaction with COX-I/II a co-crystallization with arginine was attempted with the help of Dr. Pim Puylaert. Attempts for crystallization using sila-ibuprofen were unsuccessful and only yielded oils or crystals of either of the educts, but to the best of our knowledge, the argininium-ibuprofenate structure is not yet known. A crystal was measured on an in-house-diffractometer (Bruker D8-venture, 100K, d_{min} = 1.00, for measurement details see Table 5.9). The crystal quality was limited and no high resolution dataset could be obtained. Also, the co-crystal did not contain ibuprofen carboxylate directly interacting with the arginine sidechain in the motif shown in section 1.8, but the ammonium function of the backbone of arginine. Therefore further analysis of this crystal structure was not attempted but a visualization of the final model and refinement statistics after HAR in *NoSpherA2* are shown in Figure 5.36 and Table 5.9. Refinement of anisotropic displacement parameters was unstable without restraints, therefore RIGU restraints were applied to all molecules individually. The ibuprofen is not completely enantiomerically pure. Therefore a disorder model was introduced by splitting the corresponding functional group using SplitSAME in *Olex2* and inverting the chirality of the second part by moving on screen. The disorder of the chiral center was refined to be 0.528(14) with higher presence of the R-enantiomer. The level of theory used for HAR was PBE/def2-TZVPP in

ORCA.

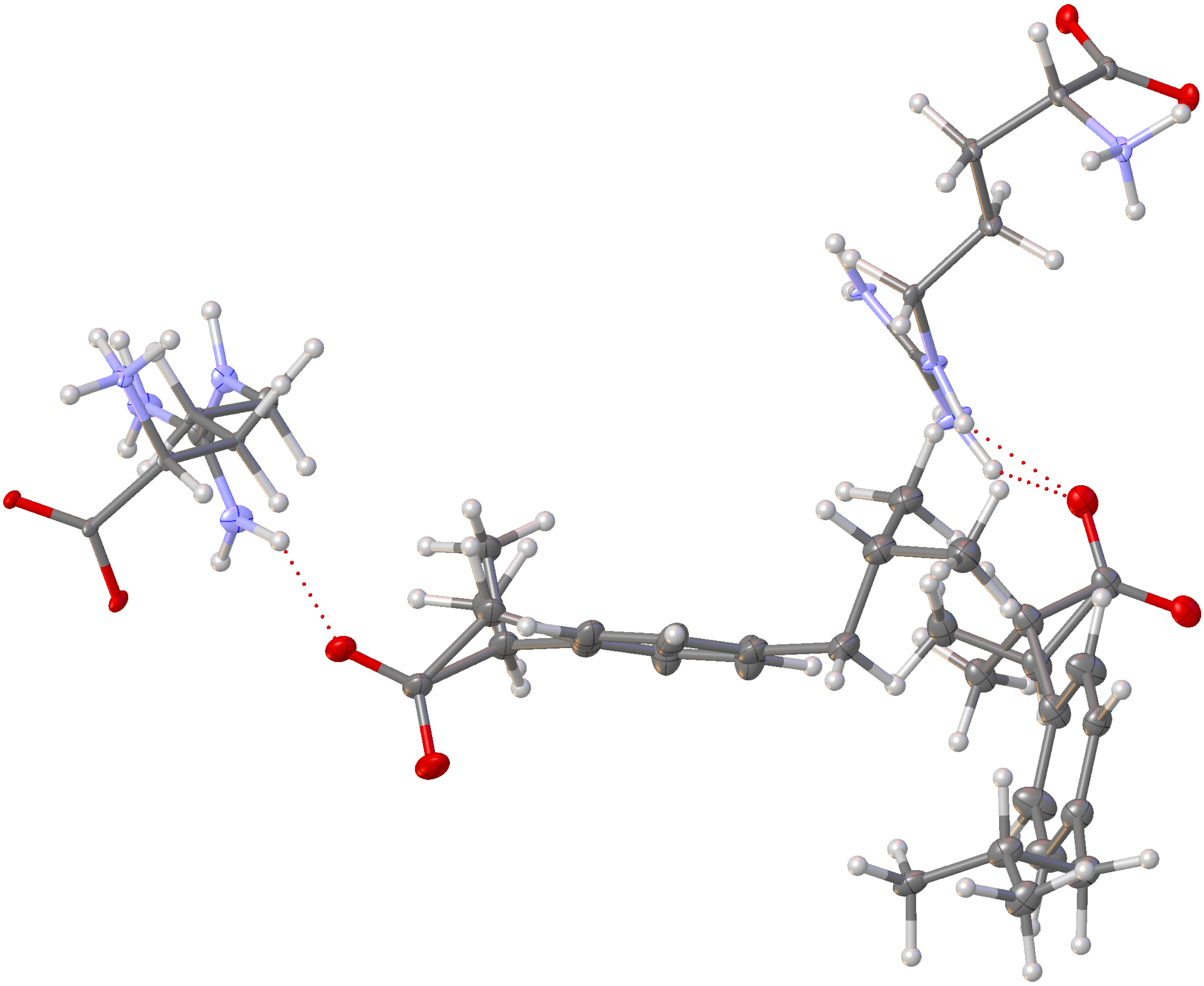


Figure 5.36: Visualization of the asymmetric unit of the ibuprofen-arginine co-crystal structure after HAR in *NoSpherA2* showing atomic displacement parameters at 50% probability level.

Table 5.9: Crystallographic measurement and refinement statistics of ibuprofen arginine co-crystal after HAR using *NoSpherA2*.

Structure	Argininium-Ibuprofenate
Space group	P $2_1 2_1 2_1$
a /Å	9.3033(6)
b /Å	14.8134(10)
c /Å	29.664(2)
V /Å^3	4088.1(5)
T /K	100
d_{min} /Å	1.00
$\lambda_{\text{X-ray}}$ /Å	0.71073
R_{int}	0.0552
Avg. redundancy	7.22
Completeness	1.00
Average I/σ	26.6
Refln. measured	51814
Unique refln.	4263
Obs. Criterium	$I_o >= 2\sigma(I_o)$
observed refln.	4097
Weighting scheme w=	$\frac{1}{\sigma^2(F_o)+(0.00P)^2+50.38P}^*$
Parameters	511
Restraints/Constraints	770/116
N_p/N_{ref}	8.35
Final R_1	0.1009
Final $R_{1,\text{all}}$	0.1039
Final wR_2	0.2110
Flack [487]$^+$	-0.1(5)
Max $\Delta\rho$ /eÅ^{-3}	0.581
Min $\Delta\rho$ /eÅ^{-3}	-0.476
CSD deposition number	

$^* P = \frac{F_o^2+2F_c^2}{3}$; $^+$ There is an error in the calculation of the Flack parameter [487] using *olex2.refine*, therefore parameters obtained using *shelx* are reported.

5.3.2 Comparison of Environments (Gas, Solvation, Crystal, Protein)

In section 5.1 (and the corresponding publication) it was shown that the crystal environment is the best possible approximation of the situation in the active site of the protein. To investigate whether this similarity of Crystal (**C**) and Protein (**P**) models can also be transferred to the crystal structures of the PEA co-crystals of ibuprofen and sila-ibuprofen, a similar setup for computations was chosen. This time the calculations were run without the use of Chemshell, [488] as a new version of *NAMD2* [471, 472] provides a direct interface configuration to *ORCA*. [54, 55] The protein references used in this study are the COX-II complexes of both molecules as produced and analyzed in section 5.2. To perform MD of the crystal system the same force field parameters were used for ibuprofen and sila-ibuprofen as for the study presented in section 5.2. The PEA molecule was modeled using a combination of parameters present in the *CHARMM*-type force field for methylammonium and phenylalanine which were combined and charges changed to result in an integer single positively charged entity while applying the fewest possible changes with regard to the original charges in *CHARMM*. The resulting parameters are given in Tables 5.10 and 5.11.

A simulation setup for model **C** of 11x11x11 unit cells, which corresponds to a cell size of ca. 65x168x245 Å for ibuprofen and 75x140x260 Å for sila-ibuprofen, was constructed based on the symmetry of the crystal using only the major component of each crystal structure and using the inverted structure of the refinement of the sila-ibuprofen PEA co-crystal, as to work with identical enantiomers for the crystal structures, that is the S-enantiomer which is biologically active one (see section 1.6 in the introduction). A visualization of the simulated crystal fragment for ibuprofen is shown in Figure 5.37. Two types of calculations were performed: A MD to obtain a measure of thermal fluctuation in the crystal to compare to the protein, and a QM/MM minimization for the calculation of a wavefunction to perform the bond-wise analysis as presented in the E64c model compound (compare section 5.1). The MD was used to calculate the aNCI. Additionally geometry optimization of ibuprofen and sila-ibuprofen anions were performed in vacuum (corresponding to the state in the gas phase, model **G**) and using an implicit solvation model of water in *ORCA*, corresponding to model (**S**). For the solvation model, the Conductor-Like Polarizable Continuum Model (CPCM) [489] of water was used as implemented in *ORCA*. The crystal and protein wavefunction originated from QM/MM minimizations.

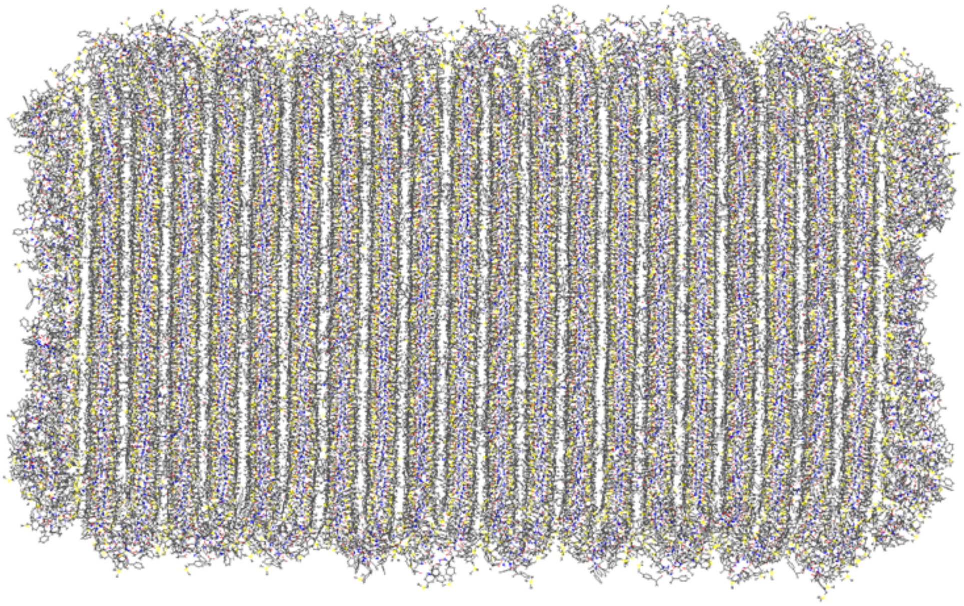


Figure 5.37: Visualization of the 11x11x11 unit cell simulation box (simulation cell size of ca. 65x168x245 Å) after equilibration used for MD and QM/MM of the ibuprofen-PEA Co-crystal.

Table 5.10: Force field parameters for bonds (top), angles (middle) and dihedral angles used for PEA.

Bond	Force constant (K_b) /kcal mol^{-1} Å^{-2}	Eq. distance (b_0) /Å
C_{T3} C_{T1}	222.5	1.538
H_A C_{T1}	309.0	1.111
H_A C_{T3}	322.0	1.111
N_{H3} H_C	403.0	1.040
N_{H3} C_{T1}	200.0	1.480
C_A C_A	305.0	1.3750
C_{T1} C_A	230.0	1.490
H_P C_A	340.0	1.080

Angle	Force constant (K_θ) /kcal mol^{-1} rad^{-2}	Eq. angle (θ_0) /°
C_A C_A C_A	40.0	120.0
C_{T1} C_A C_A	45.8	122.3
C_{T3} C_{T1} C_A	51.8	107.5
H_P C_A C_A	30.0	120.0
N_{H3} C_{T1} H_A	45.0	107.5
H_C N_{H3} H_C	44.0	109.5
H_C N_{H3} C_{T1}	30.0	109.5
H_A C_{T1} C_{T3}	34.5	110.1
H_A C_{T3} H_A	35.5	109.0
H_A C_{T3} C_{T1}	33.43	110.1
N_{H3} C_{T1} C_A	33.43	110.0
H_A C_{T1} C_A	33.0	109.5

Dihedral	Force constant (K_χ) / kcal mol^{-1} (K_ψ) / kcal mol^{-1} rad^{-2}	Multiplicity (n)	Eq. dihedral (δ) / °
C_A C_A C_A C_A	3.10	2	180.0
H_A C_{T3} C_{T1} C_A	0.04	3	0.0
X C_{T1} N_{H3} X	0.10	3	0.0
X C_{T1} C_{T3} X	0.20	3	0.0
X C_{T1} C_A X	0.00	6	0.0
X C_{T1} N_{H3} X	0.09	3	0.0
H_P C_A C_A C_A	4.20	2	180.0
H_P C_A C_A H_P	2.40	2	180.0
C_{T1} C_A C_A C_A	3.10	2	180.0
H_P C_A C_A C_{T1}	4.20	2	180.0

X is used as a wildcard, representing any atom type.

5.3.2.1 aNCI

A molecule in the middle of the cluster was chosen as the one investigated for aNCI and QM/MM. The aNCI was calculated fixing the position of the two quaternary carbon atoms of the phenyl ring of the selected molecule after the equilibration run. This way, the molecule was given the highest degree of flexibility while prohibiting translations which would compromise the calculation of the aNCI. A visualization of the aNCI of ibuprofen and sila-ibuprofen is given in Figure 5.38.

Table 5.11: Force field parameters for Lennard Jones potentials (top) and charges of atoms in e (bottom) used for PEA.

Atom type	ϵ /kcal mol^{-1}	R$_{min}$ /Å
C_A	-0.070	1.9924
C_{T1}	-0.020	2.2750
C_{T3}	-0.080	2.0600
N_{H3}	-0.200	1.8500
H_A	-0.022	1.3200
H_C	-0.046	0.2245
H_P	-0.030	1.3582

Atom type	Charge /e
C_{At}	-0.115
C_{Aq}	0.000
C_{T3}	-0.200
C_{T1}	0.100
N_{H3}	-0.250
H_A	0.090
H_C	0.330
H_P	0.115

When comparing the aNCI of ibuprofen and sila-ibuprofen in the PEA co-crystals similar patterns of interactions are observed. Dominating hydrogen bonds with neighboring amino residues around the carboxylic acid group and dispersive interactions below the phenyl ring and at the isobutyl group show that the driving force of crystal packing is dominated by these hydrogen bonds. This is similar to the interaction with the amino group of Arginine in the argininium ibuprofenate crystal structure. But since structures of both ibuprofen and sila-ibuprofen were available in the PEA system this system was used for further analysis. The amino-carboxylate N-H . . . O hydrogen bond is similar to that of the guanindino-carboxylate interaction found in the active site of COX-enzymes. The major difference is that in the PEA co-crystals the two oxygen atoms interact with two different molecules (see left side of Figure 5.38), while the two-fold hydrogen bond in the active site of COX links the ibuprofen molecules to one residue only. These three contacts on average have O-N distances in the sila-ibuprofen structure of 2.65(8) Å, 2.63(7) Å and 2.65(8) Å averaged over the whole simulation time. In the case of ibuprofen-PEA co-crystals the distances are 2.62(7) Å, 2.66(9) Å and 2.60(7) Å, respectively. To compare this the averaged values of the O-N distances for the protein environment where the guanidino function is the interaction partner are 2.77(17)Å for N_E and 2.67(8) Å for N_H in the case of ibuprofen inside COX-I, 2.66(9) Å and 2.8(2) Å in the case of COX-II, respectively. For sila-ibuprofen the corresponding distances are 2.67(9) Å, 2.72(13) Å in COX-I and 2.66(9) Å, 2.66(9) Å, in COX-II. Although some distances in the enzyme are larger and have higher standard deviations the length and strength of the hydrogen bonds are comparable, as reflected in the plots of the aNCI.

A remarkable resemblance with the plots of the aNCI of both substances in the active site of both modifications of COX is found. (compare to Figures 5.21 and 5.22) While the aNCI plot around the C-H/Si-H function is not dominating the picture the interactions between the methyl groups attached to the carbon-silicon switched position show big areas of dispersion interaction with neighbouring molecules. The hydrogen bonds in the vicinity of the carboxylate groups are the strongest interactions, as shown by their blue isosurfaces, coinciding with the observation of similar hydrogen bonds lengths. Especially in the case of ibuprofen the interaction due to dispersion around the phenyl ring is lower in comparison to sila-ibuprofen, but as discussed in the investigations of the enzyme simulations in section 5.2 this interaction is also not very pronounced in the enzyme environment. Since similar areas show interactions with the neighborhood in both systems it can be expected, that the influence on the electron density might also be similar, therefore an investigation of the interaction density is performed in

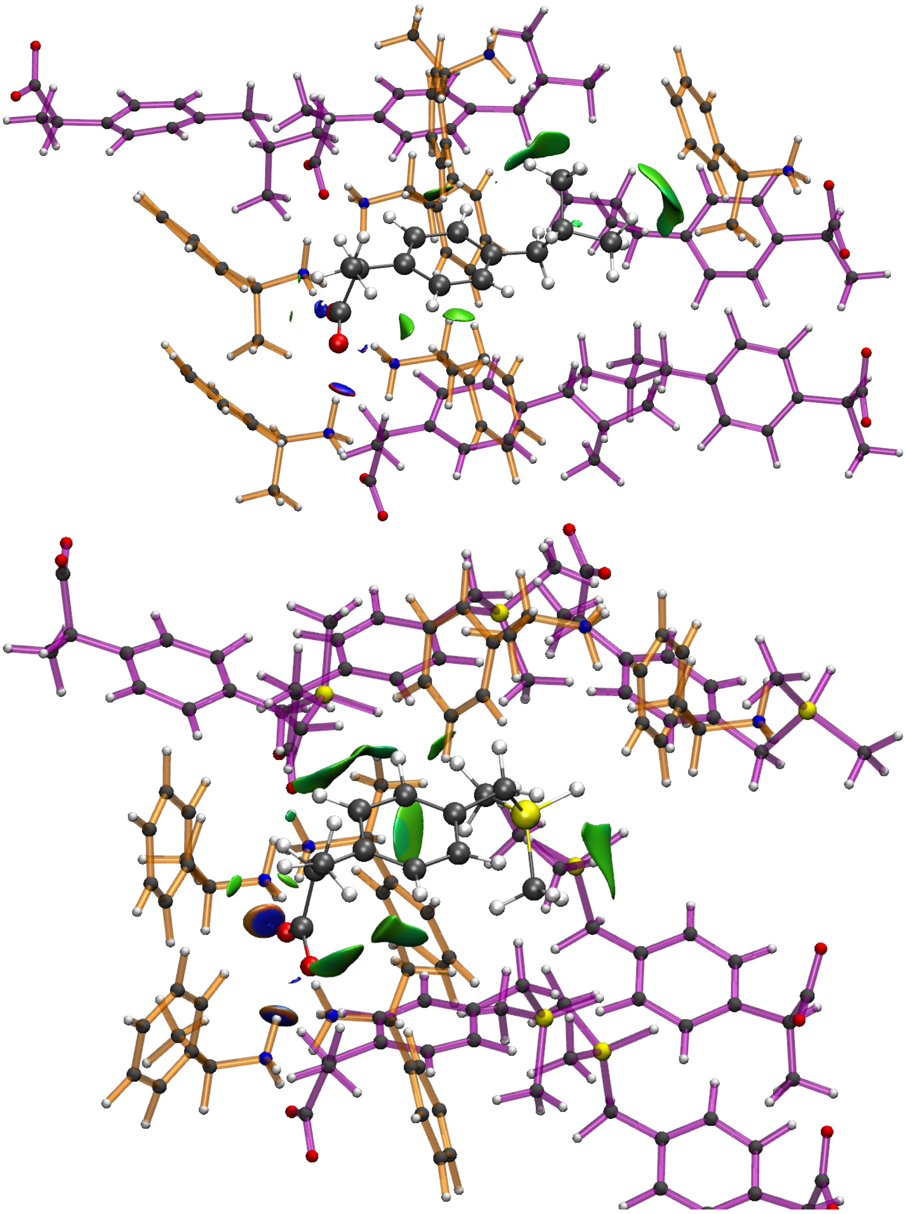


Figure 5.38: Plot of isosurfaces of the aNCI between ibuprofen (top) and sila-ibuprofen (bottom) and the neighboring molecules in the crystal. Neighboring molecules are color-coded: PEA has orange bonds; ibuprofen/sila-ibuprofen has purple bonds. Atoms were given the same color code as the main molecule. Visualized using VMD.

the following section.

5.3.2.2 Interaction Density

The interaction density is defined in the framework of this thesis (compare the study in section 5.1 and definition in section 2.3.2) as the difference between an electron density in a given environment and the electron density a molecule would express without any environmental influence on the very same geometry (e.g. without optimization). It is obtained by performing a single point wavefunction calculation on a structure that was the result of an optimization in a given environment, and subsequent computation of the electron density on a grid. This grid is chosen identically for both wavefunctions, once in the environment and once without the influence of the environment. Finally, the calculation of the difference between the grids yields the interaction density. These computations of grids were performed using *cuQCrT* (compare section 3.3). By definition, the cube without the effect, that is the one calculated from the molecule in the gas phase, was substracted from the one with the effect, that is the one in models **P**, **C** and **S**. The level of theory for the calculations used was B3LYP/def2-TZVP inside ORCA. The integrated number of electrons of the interaction density, as well as the R_{RS} are given in Table 5.12, isosurfaces of it are shown in Figure 5.39.

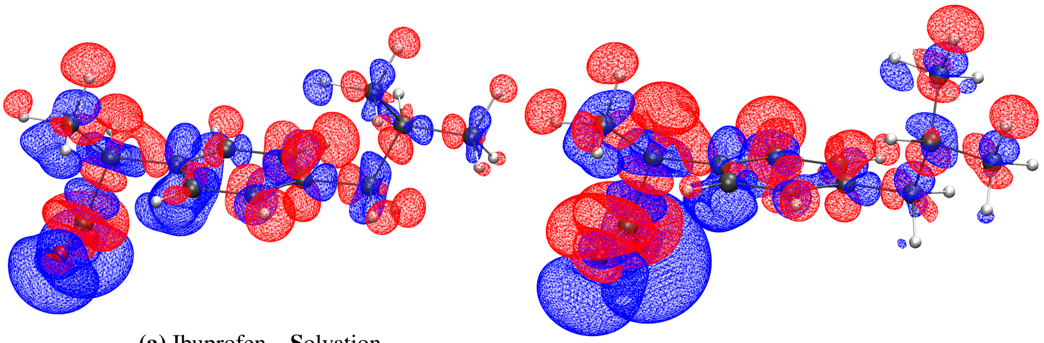
Table 5.12: Integrated number of electrons (N_e) and Real space R-value (R_{RS}) of ibuprofen and sila-ibuprofen over interaction density grids, visualized in Figure 5.39.

	Ibuprofen		Sila-ibuprofen	
Model	N_e /*e*	R_{RS} /%	N_e /*e*	R_{RS} /%
S	0.402	0.358	0.424	0.353
C	0.527	0.470	0.476	0.397
P	0.508	0.453	0.416	0.384

The order of magnitude of all interaction densities is similar in all environments. The shape of the isosurfaces is remarkably similar across both substances in all environments. Especially around regions like the carboxy group or the methyl groups, the same kind of polarization is observed. Stronger pronunciation of the lone pair regions (electron gain) or depletion of the hydrogen atoms (electron loss) can be observed. This shows that hydrogen bonding is accounted for in all environments. It is only in the carbon/silicon exchanged position that a major difference appears: in sila-ibuprofen, there seem to be only very small effects on the density distribution regarding the hydride and the silicon atom, while the ternary carbon atom is polarized in a similar way as all other carbon atoms in other positions.

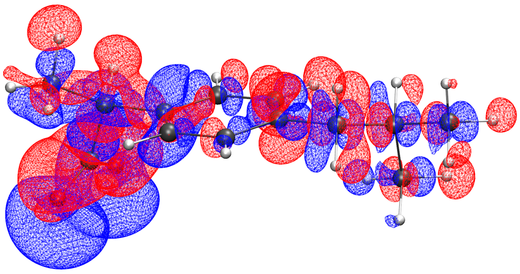
The highest total interaction density is found in the crystal environment in both cases. This is not surprising, taking into account the magnitude of electric fields of many millions of V/m needed to change the structure of molecules in crystals. [490–492] If a structural change is only achieved using these enormous magnitudes of fields, one can imagine that the effects inside the crystal have similar magnitude, which force the molecule to have the corresponding shape and polarize its electron density. Still all models have a similar magnitude of influence on the drug molecule, with the solvation model showing the smallest effect. This is reflected in the R_{RS} values as well as in the shifted number of electrons, N_e.

To comment on the differences between ibuprofen and sila-ibuprofen one has to take into account that ibuprofen has a total of 8 electrons fewer compared to sila-ibuprofen. Nevertheless it shows higher numbers of shifted electrons in all environments when compared to its silicon equivalent. This coincides with the qualitative observation in Figure 5.39 that the interaction density around the silicon atom is relatively small.

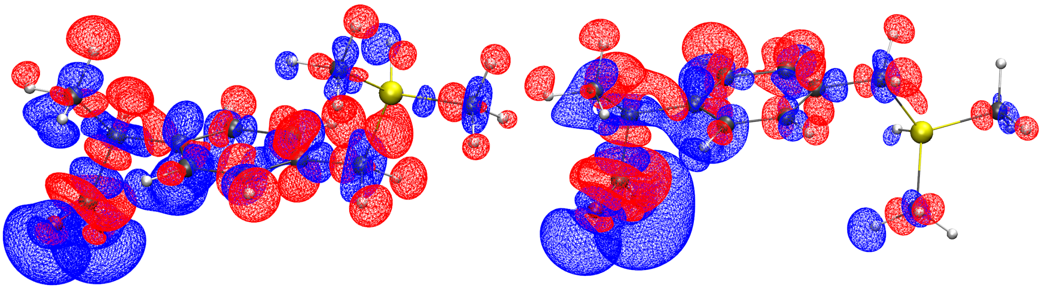


(a) Ibuprofen – **S**olvation

(b) Ibuprofen – **C**rystal

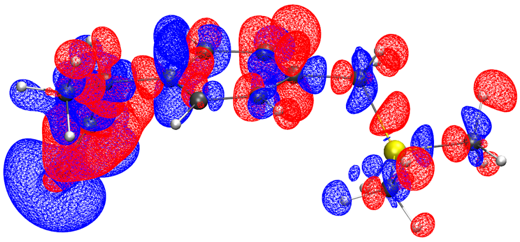


(c) Ibuprofen – **P**rotein



(d) Sila-ibuprofen – **S**olvation

(e) Sila-ibuprofen – **C**rystal



(f) Sila-ibuprofen – **P**rotein

Figure 5.39: Plots of the interaction density isosurfaces at isovalue ±0.001 a.u. (blue = positive, red = negative) for ibuprofen (top) and sila-ibuprofen (bottom) in different environments: Solvation model (a & d), Crystal QM/MM (b & e) and Protein QM/MM (c & f). Difference measures are given in Table 5.12. By definition the interaction density is **P/C/S** - **G**.

5.3.2.3 Interaction Electrostatic Potential

It was shown during the comparison of the bonding situations between ibuprofen and sila-ibuprofen in section 5.2.4 that the charges and bond types, especially in the vicinity of the silicon atom, are significantly different to the corresponding position in ibuprofen. Also, a significant difference is observed in the electrostatic potential of both substances (compare section 5.2.5). This difference in potential, where around the hydrogen atom bound to silicon even a switch of sign in the potential was observed, as well as the change of direction in the dipole moment suggests a significant difference in the response of the two drug molecules when influenced by an environment. This assumption is supported by the difference of interaction densities in the corresponding regions of the molecules as shown in the previous paragraph.

To elucidate further, the interaction electrostatic potential was calculated in the same manner as the density: Two wavefunctions – one containing the effect by the environment and a second wavefunction without any environment – were calculated and their electrostatic potential plotted in the same spatial region (e.g. grids). Then the difference was obtained by subtracting the grid of the gas phase model (**G**) calculated on the geometry from the one that actually is experiencing the environment (models **S/C/P**).

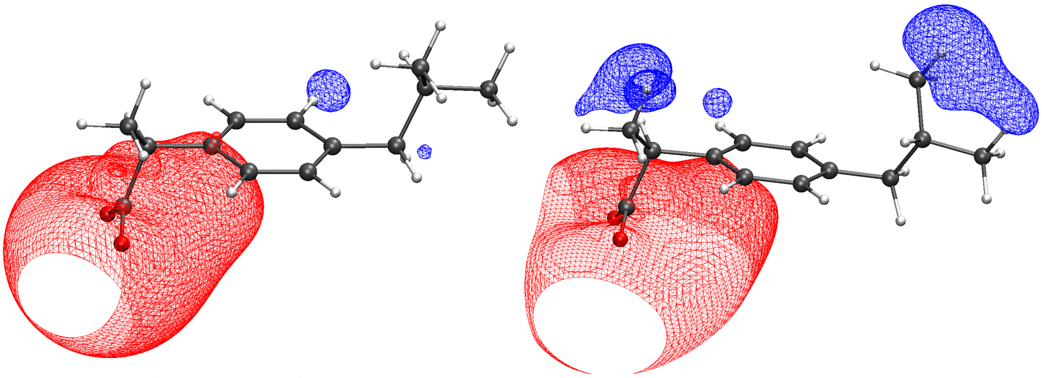
The individual plots of the interaction electrostatic potentials for both molecules are shown in Figure 5.40, the $\mathrm{R_{RS}}$ is summarized in Table 5.13.

Table 5.13: Real space R-value ($\mathrm{R_{RS}}$) of ibuprofen and sila-ibuprofen over interaction ESP grids, visualized in Figure 5.40.

	Ibuprofen	Sila-ibuprofen
Model	$\mathrm{R_{RS}}$ /%	
S	3.178	3.126
C	3.853	3.421
P	3.769	3.232

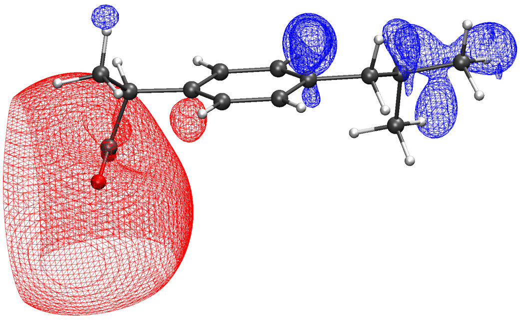
It is observed in both substances that the effect of the environment, especially around the methyl groups of the right hand-side of the molecule in Figure 5.40, is smallest in the solvation model. The crystal environment shows the highest values of $\mathrm{R_{RS}}$, while the surfaces around the aliphatic chain and the phenyl function are more pronounced in the protein systems. The higher $\mathrm{R_{RS}}$ in the crystal system is most likely due to the stronger polarization of the carboxy groups. This is reflected by the size of the respective isosurface, which around the carboxy functions is most pronounced in the case of the crystal models (see Figure 5.40).

Qualitatively, a similarity in changes of potential can be observed: the areas affected are similar between ibuprofen and sila-ibuprofen. The polarization of the potential in the protein pocket in Figures 5.40 c & f has a similar direction as the crystal and solvation models. The carboxyl group becomes more negative, reflected by blue isosurfaces, while the aliphatic/dimethyl-silyl group becomes more positively charged. This coincides with the observed pronunciation of the lone pairs of the carboxyl function in the density (compare Figure 5.39). Again, like in the density, no direct effect around the silane functional group is observed, only the neighboring methyl and methylene groups are affected by the environments directly.

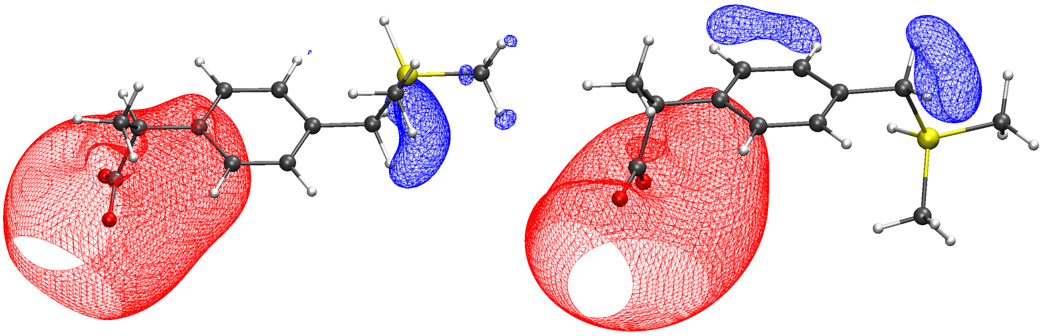


(a) Ibuprofen – **S**olvation

(b) Ibuprofen – **C**rystal

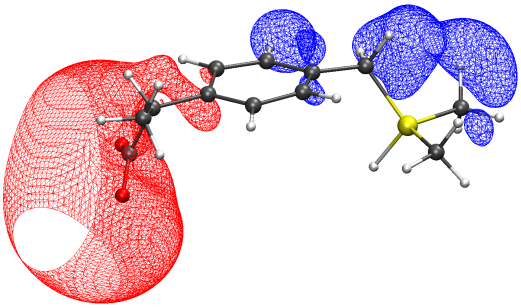


(c) Ibuprofen – **P**rotein



(d) Sila-ibuprofen – **S**olvation

(e) Sila-ibuprofen – **C**rystal



(f) Sila-ibuprofen – **P**rotein

Figure 5.40: Plots of the interaction electrostatic potential isosurfaces at isovalue ±0.028 $eÅ^{-1}$ (blue = positive, red = negative) for ibuprofen (top) and sila-ibuprofen (bottom) in different environments: Solvation model (a & d), Crystal QM/MM (b & e) and Protein QM/MM (c & f). Difference measures are given in Table 5.13. By definition the interaction electrostatic potential is **P/C/S** - **G**.

5.3.2.4 Bond-Centered Difference Density

To quantify the difference between the different environments in their native geometries, the bond-centered density calculation, which was introduced previously (see definition in section 2.3.2 and application in section 5.1), was used for all bonds in ibuprofen and sila-ibuprofen. Since the definition of calculation regions and iteration over all bonds would have been tedious and time-consuming, an automatic calculation, naming and sorting of grids was implemented in *cuQCrT* to conveniently calculate all necessary bonds with the setup of a single input file. All bonds that are found in ibuprofen and sila-ibuprofen and their respective assigned number for the following analysis are given in Figure 5.41.

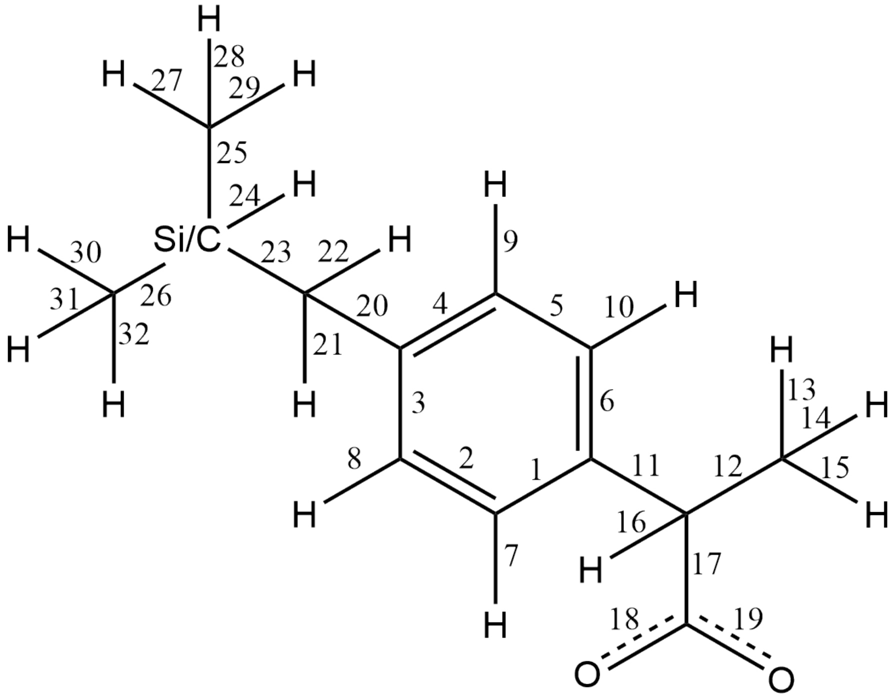


Figure 5.41: Scheme of bonds investigated in the bond-centered approach and their corresponding labels for later reference. Ibuprofen and sila-ibuprofen follow the same scheme, the C/Si switch is shown in the figure at the corresponding position.

Differences of the electron density were calculated for all models (**C**/**S**/**G**-**P**, in contrast to the previous paragraphs) and analysed in terms of the two descriptors R_{RS} and N_e. The cumulative integrated difference electron density between two environments (N_e) in each bond is shown in Figure 5.42. The cumulative R_{RS} value of all bonds is visualized in Figure 5.43. As a general remark it is noted that, even though it is possible, the R_{RS} is not reported in %, but its direct numerical value as defined in equation 2.80 is reported in the following tables and graphs.

Since the difference density of a single molecule must sum up to a value of 0 when the complete space is considered, the number of electrons shifted refers to the absolute integrated values devided by 2. The number of shifted electrons is highest in the gas phase in both substances. The solvation model reduces this difference to the protein environment, in the case of Ibuprofen by 0.7 e. In sila-ibuprofen this effect is much smaller, only reducing the difference by 0.1 e. This is interesting, since overall the differences between environments in sila-ibuprofen are much smaller than in ibuprofen.

This is especially interesting in the case of the overall shifted number of electrons, since sila-ibuprofen has additional 8 electron when compared to ibuprofen and forms more polar bonds (compare Section 1.7). One explanation of this observation might be the effect of the *umpolung*, which polarizes bonds in the sila-ibuprofen molecule internally, that are less or even unpolarized in ibuprofen in the absence of an environmental influence. The intramolecular dipole of the silicon-hydride bond might introduce a source of polarization of the surrounding bonds in sila-ibuprofen that is not present in ibuprofen. Also the difference in electronegativity between carbon

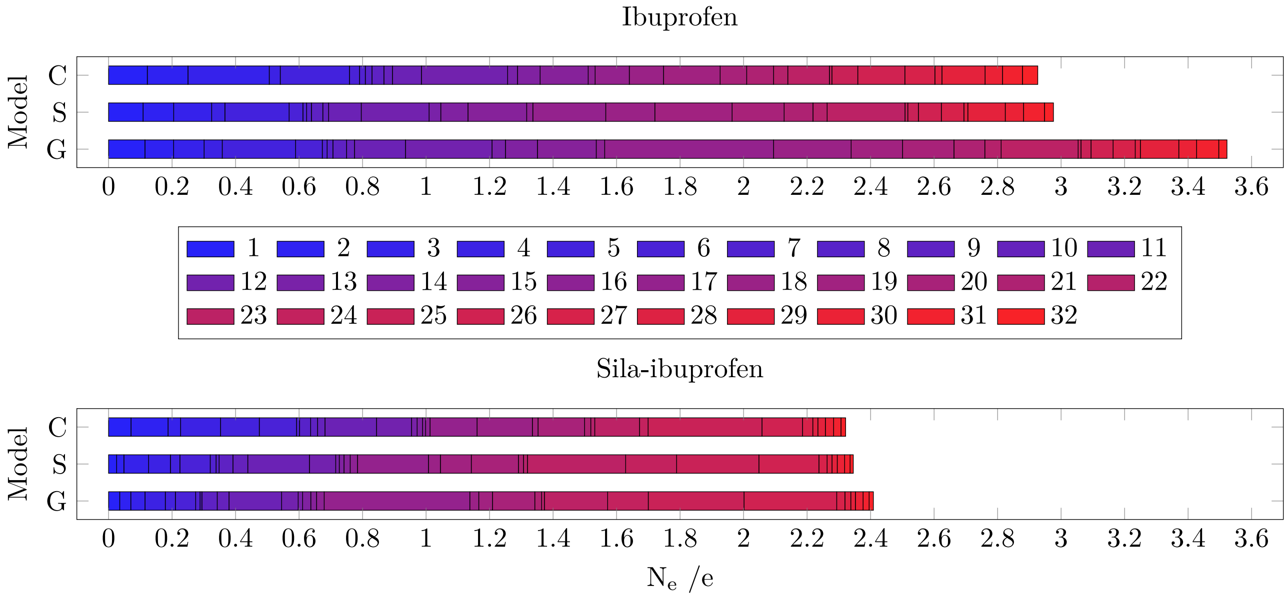


Figure 5.42: Bar-plots of the cumulative integrated difference density (N_e) in e between an environment (**C**, **S**, **G**) and the protein model (**P**) for all covalent bonds present in ibuprofen (top) and sila-ibuprofen (bottom).

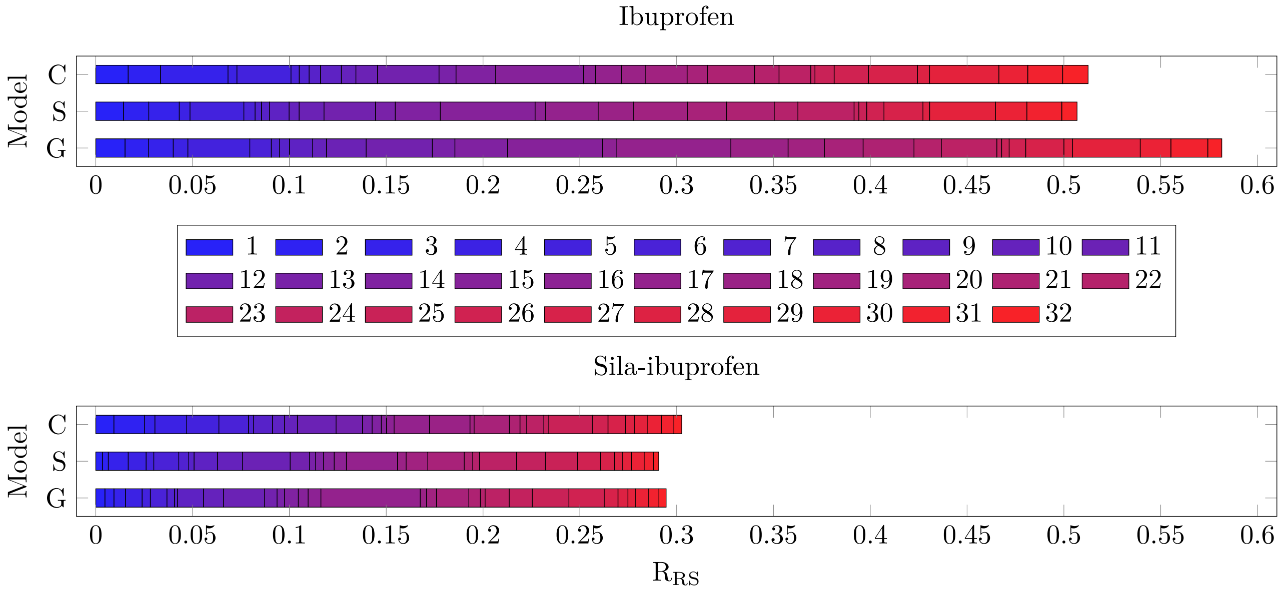


Figure 5.43: Bar-plots of the cumulative R_{RS} values for all covalent bonds present in ibuprofen (top) and sila-ibuprofen (bottom).

and silicon might polarize the surrounding carbon atoms. Since the silicon atom will donate partial electron density into the neighboring carbon atoms, it can be expected that these will also influence the neighboring hydrogen atoms in secondary effects. This would mean that bonds 20-22 and 27-32 could already show a polarization effect in the gas phase due to this intramolecular polarization. If this was the case these bonds would most likely respond less to an external influence, since the intramolecular effects persist through all environments and are probably stronger than outer environmental influences. To address this a table of QTAIM charges of the corresponding atoms is shown in Table 5.14.

The charge of the methyl group labelled C_{1,CH_3} is for example affected by the environments in a way that increases the charge in ibuprofen by approx. 0.05 e, while the change in sila-ibuprofen is between 0.012-0.025

Table 5.14: QTAIM charges in e of selected atoms around the carbon/silicon switched position of ibuprofen and sila-ibuprofen in models **G**, **S**, **C** and **P**.

	Ibuprofen				Sila-ibuprofen			
Atom	**G**	**S**	**C**	**P**	**G**	**S**	**C**	**P**
C_t/Si	0.101	0.100	0.101	0.100	2.764	2.768	2.793	2.789
$H_{C/Si}$	-0.040	-0.035	-0.033	-0.041	-0.694	-0.701	-0.704	-0.694
C_{1,CH_3}	0.004	0.058	0.054	0.061	-0.653	-0.679	-0.673	-0.667
H_{1,CH_3-1}	-0.036	-0.025	-0.011	-0.024	-0.015	-0.002	-0.015	-0.018
H_{2,CH_3-1}	-0.038	-0.023	-0.007	-0.041	-0.021	-0.003	-0.015	-0.017
H_{3,CH_3-1}	-0.042	-0.027	-0.046	-0.003	-0.022	-0.003	0.007	-0.003
C_{2,CH_3}	0.059	0.053	0.066	0.044	-0.667	-0.681	-0.678	-0.685
H_{1,CH_3-2}	-0.002	-0.023	-0.047	-0.029	0.015	-0.003	-0.010	0.014
H_{2,CH_3-2}	-0.040	-0.027	-0.040	-0.024	-0.021	-0.004	-0.005	0.004
H_{3,CH_3-2}	-0.045	-0.024	-0.010	-0.005	-0.025	-0.002	-0.019	-0.019
C_{CH_2}	0.084	0.078	0.101	0.101	-0.627	-0.617	-0.645	-0.613
H_{1,CH_2}	-0.038	-0.019	-0.041	-0.033	-0.025	-0.001	0.001	-0.015
H_{2,CH_2}	-0.042	-0.024	-0.041	-0.033	-0.028	-0.003	-0.006	-0.013
Avg. diff. to **G**	-	0.0158	0.0173	0.0186	-	0.0167	0.0167	0.0118

e. The significant magnitude of the charges in sila-ibuprofen is due to the electronegativity difference between carbon and silicon. The average absolute difference between the charge in an environment (**S**, **C** or **P**) and the gas phase (**G**) is almost always higher in ibuprofen. Only in the solvation model ibuprofen shows a slightly lower average change in charges compared to sila-ibuprofen. A remarkable difference is model **P** of sila-ibuprofen, where, on average, the charges change the least compared to model **G**.

Having this intramolecular influence of the silicon atom in mind, it is worth mentioning that exactly these surrounding bonds of silicon in sila-ibuprofen show much smaller numbers of N_e in the plot in Figure 5.42. To understand the effect of this influence of silicon, the molecule is partitioned into three different regions:

1. Bonds with silicon/carbon exchanged as a bonding partner (23-26)
2. Bonds in the vicinity of this position; one bonding partner directly bound to the carbon/silicon exchanged position (20-22; 27-32)
3. Bonds further away (remaining bonds 1-19)

1. Unique bonds for ibuprofen and sila-ibuprofen (23-26)

The plots of N_e and R_{RS} for the unique bonds around the carbon/silicon exchanged position in ibuprofen/sila-ibuprofen are shown in Figures 5.44 & 5.45. For the beginning, the focus is drawn to those bonds that make these two molecules unique.

Only looking at the N_e the difference seems to be much bigger in the case of sila-ibuprofen. The number of shifted electrons in these bonds accounts already for about a third of the total shift of electrons in the gas phase or solvation models in the case of sila-ibuprofen. In the case of ibuprofen, due to smaller numbers of shifted electrons as well as the larger number in the complete molecule, the difference is less than 15% of the total effect in the molecule. But since the total number of electrons in these grids is much different between ibuprofen and sila-ibuprofen, the R_{RS} value might give a better description of the situation at hand. This is because silicon has 8 electrons more than carbon, as mentioned earlier. In addition to the higher number of electrons inside the region, there is a difference in the charge of the silicon atom, which is in the order of magnitude of 0.025 e (compare table 5.14). This difference of charges will also be included in the grids. Since they extend further than the bond they are also included multiple times if the atom is involved in more than one bond. In this case, since

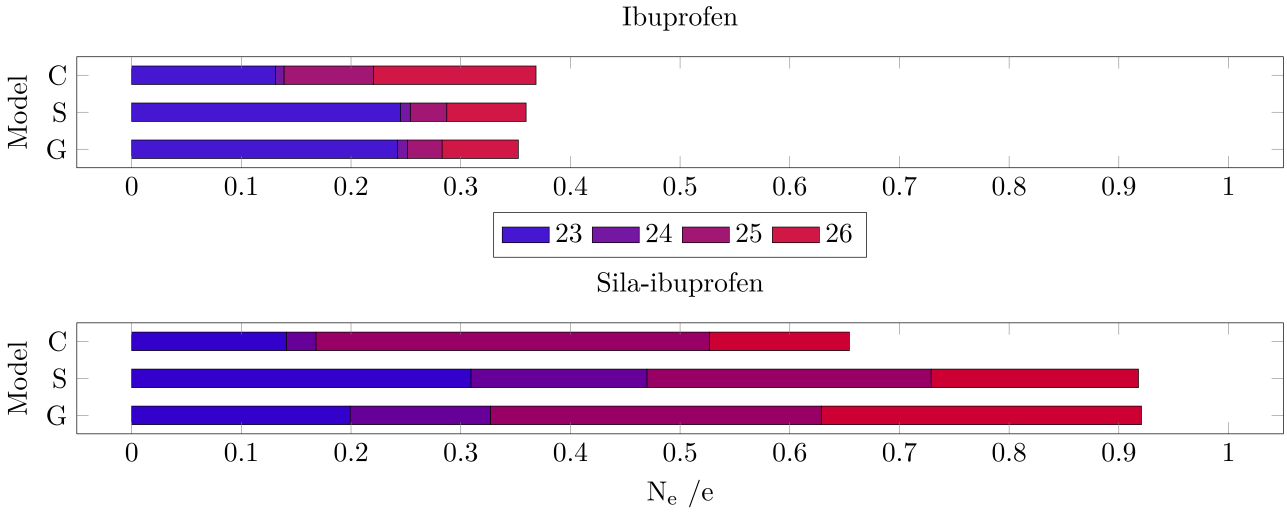


Figure 5.44: Bar-plots of the N $_e$ for the covalent bonds 23-26 of ibuprofen (top) and sila-ibuprofen (bottom).

silicon is involved in all bonds, the charge difference in silicon will be accounted for almost four times. Also, the size of cubes is significantly bigger in sila-ibuprofen, since the Si-C bond length and the C-C bond length differ by almost 30%, that is almost 0.45 Å. Since the bond-scaled method also incorporates neighboring atoms this significantly higher integrated difference electron density might also be due to inclusion of more effects of neighbouring atoms, due to the bigger box-size when calculating the absolute integral of the differences. The R_{RS} can be understood as a normalized difference measure, since the difference is divided by the local value of the density. This allows a better comparison among different elements.

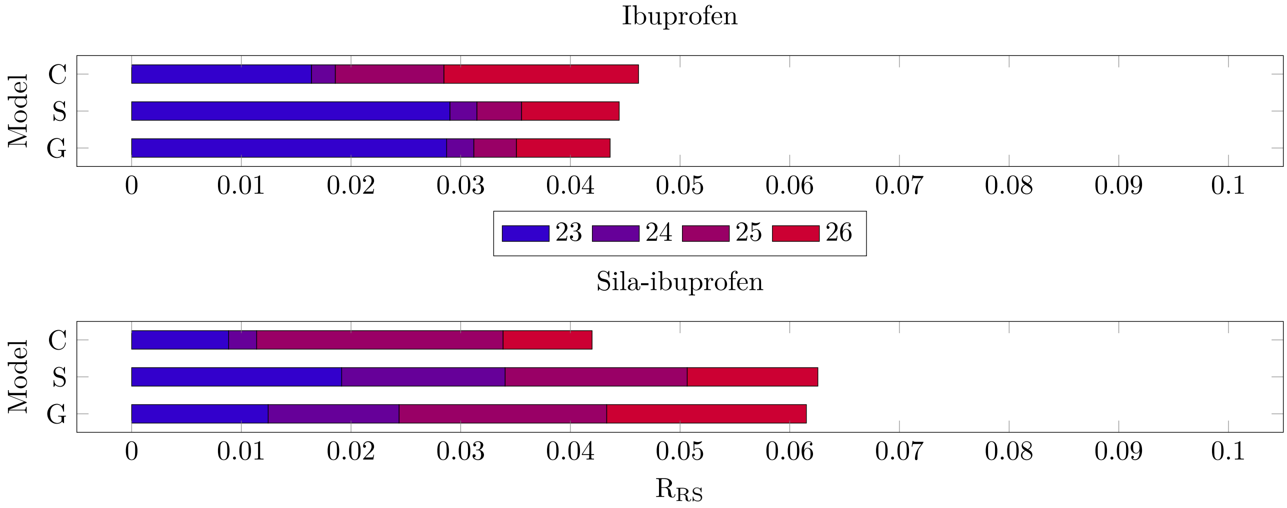


Figure 5.45: Bar-plots of the cumulative R_{RS} values for the covalent bonds 23-26 of ibuprofen (top) and sila-ibuprofen (bottom).

In comparison to the N_e where sila-ibuprofen showed around twice the values compared to the ones of ibuprofen, taking into account the value of the local density in the differences leads to a different picture. The differences are on a quite similar scale, while in the case of the crystal model the difference for sila-ibuprofen is even smaller than that of ibuprofen. While in ibuprofen, especially in the case of model S and G, bond 23 is the most polarizable, the polarization is more evenly distributed in sila-ibuprofen. Especially the polarizability of the Si-H bond 24 seems to be much higher than that of the C-H bond in ibuprofen.

2. Bonds in the vicinity of the unique bonds for ibuprofen and sila-ibuprofen (20-22; 27-32)

To investigate whether the hypothesis of less polarizable bonds in sila-ibuprofen around the silicon atom holds, the bond-wise difference for those bonds with one of the carbon atoms around the ternary carbon/silicon position are plotted in Figures 5.46 & 5.47.

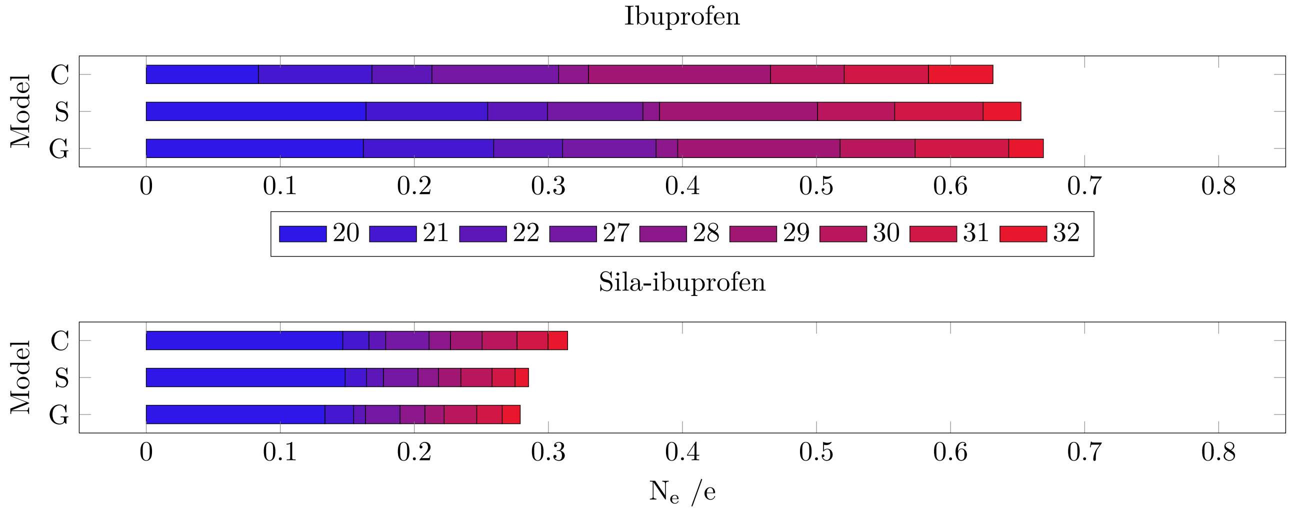


Figure 5.46: Bar-plots of the N_e for the bonds in vicinity of carbon/silicon exchange position in ibuprofen (top) and sila-ibuprofen (bottom).

A clear difference in N_e is observed between ibuprofen and sila-ibuprofen. In sila-ibuprofen N_e of all the bonds is approximately half of the corresponding bonds in ibuprofen. The highest polarization of all bonds in this set is observed for bond 20, which is the C-C bond directly next to the aromatic system. This bond is most likely affected by the delocalized ring system and therefore easier to polarize and not due to the substitution, since the effect is present in both molecules. For these bonds, the difference in total density, especially in the case of sila-ibuprofen, is much smaller compared to the first set of bonds where a different element was present. To check whether the trend is persistent when normalized for the local density the R_{RS} is shown in Figure 5.47.

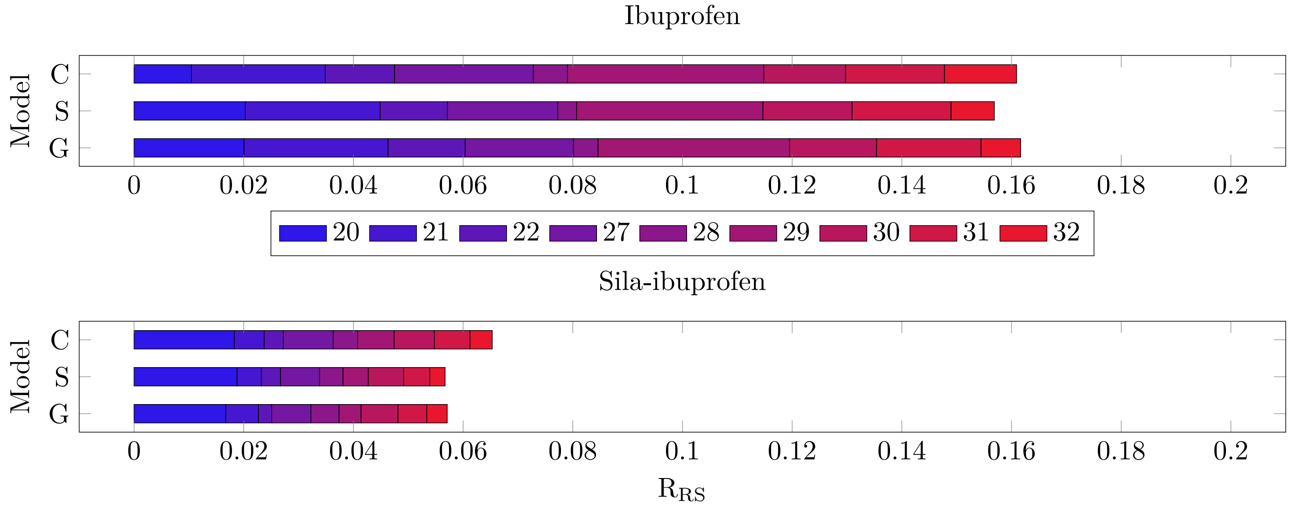


Figure 5.47: Bar-plots of the cumulative R_{RS} values for the bonds in vicinity of carbon/silicon exchange position in ibuprofen (top) and sila-ibuprofen (bottom).

In this set of bonds the R_{RS} is not as highly affected and therefore the same trend as in the plots of the N_e is observed. This means that the bonds in the vicinity of the element-switched position are in fact less polarizable in sila-ibuprofen, since the effect of the environment is very small for all bonds in all environments, especially in comparison to ibuprofen itself.

3. Remaining bonds (1-19)

To see whether the effect of the self-polarization wears off as soon as at the direct neighbors or continues further in the molecule the remaining bonds further away from the C/Si position are shown in Figures 5.48 & 5.49.

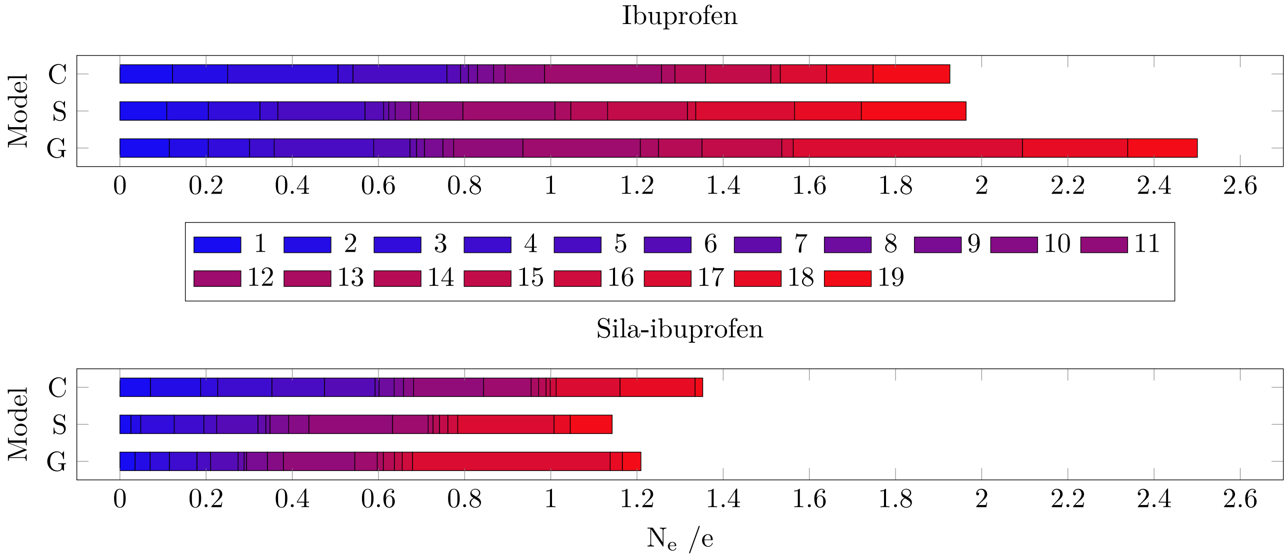


Figure 5.48: Bar-plots of the N_e for the common covalent bonds present in ibuprofen (top) and sila-ibuprofen (bottom).

Interestingly, the difference in polarization through the environments is also lower in all environments in the case of sila-ibuprofen for the remaining bonds. In the case of the gas phase comparison to the protein, the effect is only half of the polarization that ibuprofen experiences. This illustrates how the effect of the *umpolung* and elemental substitution is not only a local phenomenon but can have long-range effects. It is imaginable that this is due to the dipole present in the molecule that has long-range effects, as the Coulomb influence is proportional to r^{-2} while other effects like dispersion have a much steeper decrease with terms of r^{-6} or even lower exponents. Interestingly, the bond 17, which is the C-C bond connecting the carboxylate is the one most affected by different environments. This might be understood in terms of stabilization or destabilization of the conjugated electron system in the O-C-O system and the consequently occurring charge shift from the ternary carbon atom next to it into the system. The C-C bond becomes more and more similar to the situation found in the protein going from G over S to C. In an anti-parallel trend the sum of both C-O bonds 18 & 17 increases in sila-ibuprofen and ibuprofen. The only exception is the C model of ibuprofen, where both bonds are slightly less different than in the S model. These trends are persistent when accounting for the density, as done in the R_{RS}, as shown in Figure 5.49.

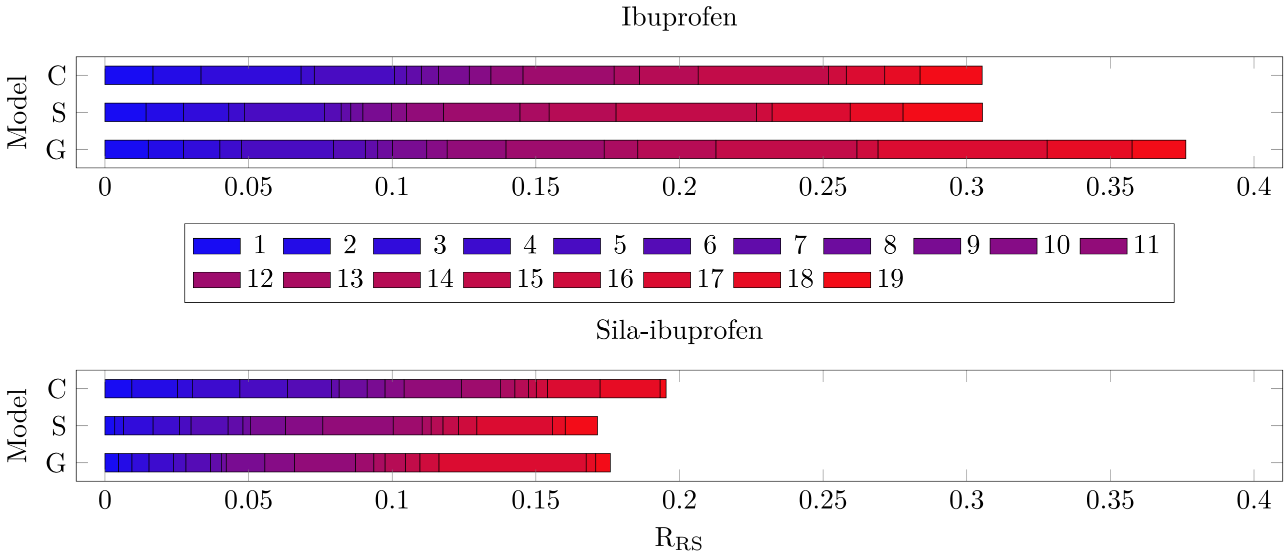


Figure 5.49: Bar-plots of the cumulative R_{RS} values for the common covalent bonds present in ibuprofen (top) and sila-ibuprofen (bottom).

In summary, sila-ibuprofen can be considered as already polarized and at the same time as less polarizable through outer influences. Therefore silicon/carbon switch, or in general elemental substitution, might present the possibility to fine-tune a molecule to withstand higher environmental effects with less response of the molecular electron density. This might provide a tool to make other known organic or metal-organic molecules less prone to polarizing effects to keep a certain shape or distribution of the density in place.
Additionally, it was shown and quantified, as is observable in Figure 5.42, that the crystal is the best possible model to mimic the density a molecule will express in the environment of the active site of the protein it targets in the biological media, as has been shown in section 5.1. This confirms the idea of not only structural but also density and molecular properties being best understood from a crystal structure to make predictions for the situation in drug applications, as proposed by *Klebe*, *Pascard* and *Luger*. [274, 275, 282, 493]

Chapter 6

Summary and Conclusions

The main objective of the work leading to this thesis was to develop new or improved crystallographic methods and to apply them to quantum crystallography and drug design.
The methodology (and the software) produced during this work leads to a better understanding of the molecular structure based on a set of improved procedures: i) better atom positions and other crystallographic parameters from advanced structural refinements of X-ray diffraction data ii) better methodology to describe similarities and differences of molecular electronic structure in various environments (based on tailor-made grids of properties like densities and potentials for bond-centered comparison between different conformations and environments) iii) new models to incorporate thermal fluctuations and movement in the analysis of interactions between molecules using the aNCI and aIE.

These new or improved functionalities were performed making use of modern, either GPU accelerated or, at least, fully parallelized shared memory, code. This facilitates detailed studies of interactions between atoms and molecules, even taking into account the evolution of interactions through time by a combination of otherwise purely static descriptors and molecular dynamics simulations. This new software, called *cuQCrT*, has also in many parts become the engine allowing the fast calculation of scattering factors in *NoSpherA2*. The interface for Non-spherical atoms in *Olex2* (NoSpherA2) was developed through a close collaboration with the developers of *Olex2*, one of the most common software used for modern small molecule X-ray diffraction refinements.
We can now provide a standardized interface for Hirshfeld Atom Refinement and any other atom-based non-spherical refinement method to be used in a convenient way using and well-established refinement software. Using this platform also offers various additional benefits to analyse the results of HAR, for example plots of deformation densities on screen as you refine, wavefunction derived properties like potentials, reduced density gradients or the Laplacian, as well as electron localization to be plotted on screen. In addition, this generalization enables the direct and unbiased comparison of different non-spherical refinement methods for the first time, since the same refinement engine and settings can be used readily. The re-implementation of Hirshfeld stockholder partitioning in *NoSpherA2* allows the comparison with the previous implementation (in *Tonto*) in detail and provides information that is required for the development of further and more advanced methods.
Major improvements include the fact that disordered models can now be processed (which is crucial for any biological application) and that structures in *all* space groups, including high-symmetry space groups with many atoms on special positions, which could not be run in *Tonto*, can now be examined. The latter is crucial for the refinement of the type of samples that are important in materials science.
The inclusion of routine treatment of heavier elements, as well as the possibility to use open-shell and high-spin wavefunctions, opens the quantum crystallographic method development to the community of organo-metallic scientists. This advance in treatment of heavier elements and flexible electronic structures of metal-organics is mainly due to the implementation of an interface to routine QM software like *ORCA*, *Gaussian* or *pySCF*. Since this software is already available, in the case of *pySCF* even in the form of open source code, the standardized calculations and subsequent use of wavefunctions is much simpler compared to implementations in *Tonto*. This is further improved by the convenient availability of relativistic calculations when using routine quantum chemical calculation engines, as those mentioned. Additionally, the availability of an easy to use graphical user interface (GUI) in *Olex2* makes non-spherical advanced structural refinement – and ultimately the field of quantum crystallography – available to the majority of scientists that work using single-crystal diffraction data.

In addition to improved structural description based on advanced structure refinements tools developed during this thesis allow the careful analysis of wavefunction fittings, as performed in section 4.1. After modelling the structure to the best of our knowledge based on structure factors based on purely theoretical wavefunctions the wavefunction fitting represents the fitting of the charge density in a quantum mechanical way. These investigations show that in principle the XCW fitting method is capable of retrieving various effects from diffraction data, including polarization and electron correlation. Also it was shown that for urea experimental data can provide sufficient information on the effects like polarization and electron correlation to improve a wavefunction using XCW fitting, even if a more sophisticated method is chosen. This was further confirmed by including some of the effects in the quantum mechanical *ansatz* and observing a qualitative similar shape of the difference densities of these purely theoretical electron densities compared to the fitted wavefunctions using an inferior level of theory. It was also shown that a method that partially or even in exaggerated amounts contains these effects effect, can be improved by XCW fitting, as it was the case for DFT and electron correlation. The DFT functional showed too much electron correlation and the agreement to CCSD reference calculations could be improved upon inclusion of information through structure factors. In turn this allows the retrieval of polarization and correlation not modeled by the *ansatz*. A scheme to theoretically divide observed changes of the density into contributions of different effects like electron correlation and polarization by use of combinations of individual fittings with different effects included *a priori* in the model chemistry was proposed. It qualitatively shows that a separation of physical effects of electron correlation and polarization, even when using experimental data, is possible. It remains to be investigated whether this points towards a fundamental complete separability of these effects.

These observations have implications for HAR, as then a level of theory should be chosen to model the electron density as accurately as possible in order to obtain the best structure from advanced structural refinements before an attempt is made to retrieve the effects from the experimental data.

The investigation of interaction densities in the E64c model compound (see section 5.1) shows that the effects on the electron density of a drug molecule can be modelled best by the situation a molecule experiences in a crystal. Neither solvation nor gas phase model – even with explicit inclusion of a counter ion – were capable to reproduce the density in the protein environment as good as the crystal structure does. In the case of HAR, especially when cluster charges were used, the density of the obtained wavefunction might be a good guess for the situation found inside the active site of a protein and therefore could be proposed as a new tool to derive active site binding matches based on database approaches like *IsoStar* etc.

Also during this study it was shown that if diffraction data is used that is of inferior quality, as it in this case contained radiation damage effects, the fitting procedure will highly be affected by these undesired effects. This could, in principle, be the case for incorrect or insufficient absorbtion correction, strong fourier-truncation errors and other effects, as well. Therefore great care must be taken when using experimental data. In experimental data one can never blindly trust, to counter the statement made by Prof. Stalke *et al.* [494].

The reconstruction of the crystal field experienced by molecules in the crystalline phase using XWR is going into the right direction, as shown by the plots of percentage reconstruction in Figure 4.13. But it has to be taken into account that the polarizations, especially in terms of atomic charges, is also very sensitive to the quality of experimental data. Any defect in the low-order reflections, which are shown to be the most important information for the modelling of the valence and differences from a spherical promolecular density (compare section 1.6), can introduce large discrepancies into the description of the molecular density. This was especially severe in the case of the E64c model compound (section 5.1).

We were able to show that XWR can retrieve correctly the bonding features of a molecule from diffraction experiments conducted on a lab-source, which pure DFT was unable to predict. The study of the ethylene-push-pull systems in section 4.2 shows the possibility to model the electronic structure and non-classical bonding features of a molecule from in-house diffraction experiments – if sufficient data quality is available. Especially since the X-ray beams of many large-scale facilities reach intensities that have a significant, undesired impact on the crystal in inelastic ways, e.g. radiation damage, the more widely available and cheaper in-house-sources might become the device of choice for a routine conduction of quantum crystallographic investigations.

We showed the synthesis, characterization and quantum crystallography of sila-ibuprofen, the carbon-silicon switched equivalent of ibuprofen. The procedure of XWR was used to develop force fields based on wavefunctions obtained from diffraction data restrained calculations for this purpose. It was tested for toxicity and *in-vitro* activity as a new potential NSAID. We successfully showed that the presented new methods for analysis of molecular interactions, namely aNCI and aIE in combination with the established FEP calculations, can predict the similarity of a novel drug molecule *a priori* in the case of the ibuprofen derivative sila-ibuprofen.

An investigation of the similarity between environments of a molecule in the crystal structures of co-crystals of ibuprofen and sila-ibuprofen with PEA and other environments like the target enzyme or solvation confirmed the conclusion, that the crystal system and the effects of polarizations inside the crystal are in fact a good estimate of the electron density of a drug inside the active site of a protein, as it was already found in the case of the E64c model compound. These investigations also conclude that an implicit solvation model is already a much better estimation of effects than assuming a single molecule wavefunction in the gas phase. The consistent observation of this trend has interesting implications for the calculation of wavefunctions for Hirshfeld Atom Refinement, as this might indicate that the use of these solvation models might improve the estimation of the atomic form factor without the use of an explicit clusters of molecules or the expensive optimization of cluster charges to mimic the crystal field.

The investigation of the similarities and differences of environmental effects on ibuprofen and sila-ibuprofen showed the effect that sila-substitution has on the response of an organic molecule to the influence by the environment. The dipole of the silane functional group allowed internal polarization which led to a decreased response to environmental effects. This might be an attractive motif and property for further drug development.

One of the most fundamental chemical findings during this work was the unambiguous determination – from experimental data – that the non-spherical description of a so-called ionic salt (like CaF_2) is in much better agreement with measured structure factors by X-ray diffraction than the classical spherical ion – or in the case of the IAM even neutral atom – model. This points towards a directed bonded interaction, which completely refutes the definition of ionic salts being formed by Coulomb forces alone. These findings challenge our textbook notion of rock-salts and the description of single atom ions in general, as long as they are not isolated in the gas phase.

Chapter 7

Perspectives

The presented interface *NoSpherA2* will allow an application of quantum crystallographic methods for many fields and in an intuitive way, allowing also non-experts to perform these refinements. This will hopefully in a long term improve our understanding of molecular structures, not exclusively limited to the crystalline phase. It also will allow the standardized use of a common interface and therefore make different file formats and input structures, e.g. all the different flavors of multipole-based techniques, obsolete. Hence it allows the use of databank techniques, like for example DISCaMB/MATTS, [80, 81] within the normal framework of refinement used by most crystallographers.

Since the refinement of structures is not only limited to the use of X-ray diffraction experiments, an extension to allow the refinement of electron diffraction data (see [462]) based on the HAR methods will be applicable and within reach. Collaborations aiming to solve this issue are already starting. Ultimately, one might also think about a joined refinement using both X-ray and electron diffraction data in order to retrieve different aspects that are more pronounced in either of the methods.

NoSpherA2 also allows the visualization of various properties and descriptors for a deeper understanding of a molecule under investigation. These possibilities and features are to be extended to allow a sophisticated quantum crystallographic complementary bonding analysis based on the crystal structure at hand.

Since the implementation of *NoSpherA2* allows use of routine QM software, it can now be easily evaluated what the effect of various levels of theory are on the quality of a structural refinement within reasonable timescales. This will allow a systematic study to find the most suited level of theory – including basis sets, method and integration accuracy – to perform HAR. Also effects of polarization techniques like solvation models, the implementation of cluster charges similar to the ones used in *Tonto*, post-HF methods like correlated wavefunction calculations and the use of periodic boundary conditions might be explored for their advantages or disadvantages in HAR.

It remains to be evaluated if the use of the averaged interaction energies (aIE) contain additional information like the entropic terms of e.g. lattice vibrations, which could be evaluated from the fluctuation of the different energy terms. But to obtain a better description of the difference in binding in the active site of a protein, the implementation of the calculation of the aIE in a solvated state to reference the protein bound situation against might provide a more accurate prediction of differences in binding energy between different drug molecules.

A recent publication [151] reporting excellent predictions of electron densities of organic molecules through machine learning promises accurate electron density distributions in very short time compared to QM calculations. This approach could provide instantaneous densities to be used for the Hirshfeld partitioning and could make HAR a procedure that could become fast enough to be useful for practical applications. A different route to improved speed is the use of ELMOs for the calculation of a wavefunction. This is already reported for HAR in *Tonto* and will also be implemented in *NoSpherA2*. This might allow advanced structural refinement results live during a data collection processes with the sophisticated information of wavefunctions being present. Since the data treatment of X-ray diffraction data is still mainly focused on the IAM, but absorption and radiation damage processes have an impact both on the core excitation and later through the relaxation in the valence region of atoms, the non-spherical description might provide deeper insight and ultimately better data treatment.

The analysis of the disagreement of structure factors in HAR might allow the conclusion that the Lagrange parameter λ in XCW should not simply introduce all disagreement statistics between experiment and model, but maybe select significantly disagreeing reflections and describe these first. If all diffraction data is introduced at the same time the relative importance of each individual disagreeing reflection is reduced by all high order reflections, which usually agree quite well, even in the IAM. Effectively this slows down the inclusion of all effects observable in the data, as shown in section 4.1. The application of significance dependent inclusion of reflections for the restrained wavefunction calculation based on diffraction data could be implemented and intrinsically allow a simultaneous monitoring whether the constrained wavefunctions starts over-fitting the diffraction data, since the set of previously agreeing reflections should not start to disagree between fitted wavefunction and diffraction experiment during the refinement.

Sila-ibuprofen was proven to be as potent as ibuprofen in theoretical models and *in-vitro* experiments while having favorable physical properties. The next stage for the approval as a drug would be animal testing to obtain parmacokinetical data and ultimately get to clinical trials. This was, however, far beyond the scope of this thesis and remains to be done by experts in that field. An application of an organo-silicon drug would be an interesting setting, since carbon and silicon are in the same group of the periodic table but life on earth is only based on carbon, while other element of the same group are used (see O, S and N, P) within living organisms. It is also of interest whether the human body is capable of forming the metabolite of sila-ibuprofen in the same way as for ibuprofen. Since ibuprofen is usually oxidized by the human body at the position where the carbon silicon exchange was performed it would be interesting to investigate what metabolites of sila-ibuprofen are observed. This is especially interesting in the context of current experiments showing an evoluted version of P450, the enzyme responsible for oxidation of ibuprofen in the human body, [299] that is capable of oxidizing silanes. [495]

As new research questions arise, scientists need to grow, change and adapt their methods to succeed in answering these questions. Method development – exemplified here by the development of new and modern software – placed somewhere at the interface of chemistry, crystallography, computational electronic structure modeling, biochemistry, and programming (underlined by the collaboration with many co-authors) is crucial to keep the ever-spinning motor of research running. Many people may think of crystallography as a fully-matured field which is mainly used as an analytical tool for the reliable determination of atomic connectivity. But this thesis shows beyond any doubt the benefits that various fields of science can draw from these newly developed quantum-crystallographic methods.

In this work, not only could we harness the power of Quantum Crystallography to experimentally and unambiguously answer a long-standing open question about non-spherical ions in rock-salts or the zwitterionic character of twisted ethylenes, but we could even present a new drug molecule based on purely rational design and predict its potency. In this way, these newly developed quantum crystallographic methods – alongside the recently developed foundations of the field – promise new possibilities in the analysis and fundamental understanding of chemical bonding an weak chemical interactions. Moreover, these advanced structural analyses using diffraction experiments ultimately lead to better rational development of drugs and materials.

This might pave the way towards the fully automatized analysis of new (drug) molecules by diffraction experiments and subsequent quantum crystallographic analyses. The future evaluation of a drug candidate by the crystal phase will not only be substantially faster than ever before but invaluable information relating to the pharmaceutical potential of the proto-drug will be routinely obtained at an early stage. The future development of exciting new methods may well be driven by new and emerging fields in other sciences, which might contribute to Quantum Crystallography just as the work presented in this thesis will contribute to other scientific fields: The development of scientific methods is at the very heart of all science.

As *Friedrich Nietzsche* wrote in his book "Human, all too human":

"On the whole, scientific methods are at least as important a product of research as any other result: for the scientific spirit is based on insights into method, and if those methods were lost, all the results of science could not prevent a renewed prevalence of superstition and nonsense."

Acknowledgements

First and foremost I want to thank my thesis supervisor **PD Dr. Simon Grabowsky**. You taught me not only how to become a scientist, but also allowed me to explore and experiment with various ideas during this work and helped me when my ideas became to vague and abstract to stay on track and focused to finish the projects. I am very grateful for your guidance and the possibility to experience working in your group and developing these exciting and fascinating things.

I also want to thank **Prof. Dr. Jens Beckmann** for letting me work in his labs and the nice collaborations. Not only did you lend me your equipment and chemicals but also agreed to be my third assessor for my PhD defence, thank you very much! Additionally I want to thank **Prof. Dr. Chérif F. Matta** who agreed to be the assessor for my written thesis. It is an honour to present my work to you and to receive feedback from such a well-known and renowned quantum-biochemist. Thank you very much for the kind correspondence and your agreement to have a look at what we have been working on. I also want to thank you for your time and the extremely helpful suggestions to make the best of my thesis. In addition I want to thank **Prof. Dr. Matthias Arenz** for agreeing to be the chairman for my PhD defence.

The Funding for these projects is greatly acknowledged. The Sila-Ibuprofen study was funded by the *Studienstiftung des deutschen Volkes* and the *Forschungskommission* of the University of Bremen.

How boring and quiet would it have been in the office without my colleagues. Thank you guys! **Dr. Malte Fugel** already had to deal with me being his group-mate since the Bachelor Thesis. Thank you for many discussions and the great time we had in Japan during the beamtime at SPring-8 and in Tokyo! I also want to thank **Dr. Lorraine A. Malaspina** for being my office- and group-mate and an amazing opponent for heated discussions. It would have been way too quite in the office if we would not have had each other. Also thank you for the great BBQ-partys at your place and always being a critical discussion partner to question my crazy ideas. **Daniel Duvinage**, you have helped me countless times in the lab, especially when my two left hands could not do what I wanted them to do. I guess if it wasn't for you I would have never had any material to work on, to begin with. Thank you for the many funny memes, breaks and work-hours we spent during my time at the Beckmann's. In addition I want to tell **Dr. Pim Puylaert** that I am deeply sorry for asking you to help me with the crystallization of sila-ibuprofen co-crystals. I am amazed how many attempts you took to obtain these results, even after I left Bremen. Thank you also for many discussion about the normal things besides chemistry in life!

I also want to thank the students in the AG Grabowsky in Bremen for their help and the nice atmosphere we had in the group! Thank you **Erna Wieduwilt** for various discussions about life itself and humans and their behavior! Thank you **Anneke Dittmer** for being a very nice office-mate and helping me to learn ORCA, which you also just learned. After Anneke left **Justin Bergmann** came to our office and has been a very pleasant colleague to have. Thank you for the workout-challenges we had and the nice discussions! Last but definitely not least of the *CrystalWarriors* I want to acknowledge **Dr. Emanuel Hupf**, who taught me how to read a *.cube* file in the beginning and gave me his code which basically started all the mania about the bond-wise analysis! You taught me many things about computers in the beginning of my work in the group and were an amazing host for the BBQ and s'mores at your new house! That was awesome!

Since I moved to Bern **Dr. Michelle Ernst** was an awesome office-mate! Thank you for your patience to show me how thinks work in Bern and Switzerland and at the Departement für Chemie und Biochemie (DCB)! It was really nice to share the office with you and always have an open ear to complain to if something went wrong! You also gave me incredibly helpful and constructive feedback after you read through all of my introduction and theory of this thesis! Thank your for this great time and I hope we can continue to stay in touch and maybe also arrange a scientific collaboration in the future! I also want to thank **Dr. Michał Andrzejewski** and **Dr. Jürg**

Hauser for the warm welcome and showing me around the crystallographic and IT equipment at DCB. I am still amazed how organized and well working things are since Day 1.

In my opinion a scientist should always have a senior scientist in their environment to critically ask: "Well, what do you expect?" or "Great, but what is the point of all of this?". Thank you **Prof. Dr. Hans-Beat Bürgi** for always having an open ear to my thoughts and ideas, regardless of their usefulness. Your endless interest for current developments and your broad overview of the crystallographic field of science inspires me and shows me that being a scientist is more than a title or a job. **Prof. Dylan Jayatilaka** introduced me to the impressive but also sometimes frustrating world of scientific and quantum-crystallographic programming and his "swiss-army knife of Quantum-Crystallography". If it wasn't for your inspiring work and software basically nothing of what I did would have been possible. Thank your for introducing me to your code and agreeing to let me have a try at programming it. I also want to acknowledge **Prof. Dr. Alessandro Genoni**, whose code to translate *.wfn* file format MO-coefficients into *.fchk* style and order, as well to calculate density matrices fueled my attempts to write a software to handle wavefunctions. This basically laid the kick-off point on my way to write *cuQCrT*, which ultimately became part of *NoSpherA2*.

It was, and still is, a great pleasure to work with the *Olex2*-Team. Thank you **Dr. Horst Puschmann** for already introducing me to programming inside *Olex2* already back in Hyderabad. I also can't stress how much it helped me that you read my thesis and went through the hassle to work on corrections with github! **Dr. Michael Bodensteiner** talked me into working with you guys which I was sceptical about in the beginning, since I feared to not be skilled enough. But the *Olex2* people were very welcoming and made a great effort to incorporate me and left me free hand to design *NoSpehrA2*, which was an incredible gift from you! Therefore I am also in great debt of **Dr. Oleg V. Dolomanov**, who created an awesome backend and engine and was always listening when I thought I found another bug in the code, which was just a clumsy user in front of the screen! It was, and still is, amazing to work with you guys!

When I moved to Switzerland I feared to have no time to make more friends. **Georgia**, **Giuseppe**, **Fulvio**, **René**, **Tomasz**, **Lars** and **Giovanni** proved me wrong, since they became an awesome group to play boardgames, have a BBQ on the weekend in the idyllic forests of Bern or have ice-cream at the Gellateria di Berna and dicuss until the middle of the night. You made my stay in Bern very nice and I am very grateful for all the time we spent, while we could all meet.

None of my programming and IT skills would have been possible without my father **Holger Kleemiß**, to begin with. You taught me to use a computer before I could read and helped me understand more complex programs and programming ideas already since school-time. I thank you very much for your help, endless discussions during weekends and lunches & BBQs at home and for teaching me almost every time anew that I know nothing about programming!

Last, but surely not least, I want to thank you, **Stella Geiger**, for being there. For being a part of my life and for being so understanding with my world I sometimes live in. You encouraged me to take the step to move to Bern and stayed with me the whole time. You had my back when I needed support and were there for me when my mood swings got me. You are exactly the right combination of joking and philosophical, curious and relaxed, culturally active and fine with gaming together. I am in great debt for you hosting me for such a long time during this whole pandemic which is happening this year and for still reading my thesis even though I never stop talking about my work, anyway! I hope we will also have great times in the future and find a place to stay.

Florian Kleemiss

Appendices

Preface to the Appendix

This Appendix contains a chapter with the Software Reference (chapter A) and the additional information regarding chapters 3, 4 and 5. The Supporting information is printed as deposited with the journals for the corresponding publication with identical name and commented with a paragraph summarizing the individual contributions made during the work on this thesis. Other appended material, like software code, is not commented, since it was referred to in the main text.

Appendix A

Software Reference

A.1 AIMAll

AIMAll version 19.10.12 was used in this thesis. [443] Also the visualization tool AIMStudio of this software was used to conveniently analyse the results through the GUI.

A.2 Crystal

The program package *Crystal14* was used for periodic or molecular calculation of the wavefunctions and densities.

A.3 CrystalExplorer

CrystalExplorer version version 17.5 was used in the course of this thesis. [423]

A.4 DGrid

DGrid Version 5.0 and after its release 5.1 were used for the ELI and Raub-Jansen Index calcualtions during this thesis. [441]

A.5 Gaussian09

Gaussian09 revision D was used wherever Gaussian is mentioned in this thesis. [438]

A.6 Multiwfn

Multiwfn was used for plotting properties on lines between two atoms or for different transformation of files. [439]

A.7 NAMD2

NAMD2 was used for molecular mechanics, molecular dynamics and QM/MM calculations during this thesis. [471,472]

A.8 NBO

NBO version 6.0 was used for the NBO and NRT analysis during this thesis. [445] Input files were generated using QM packages like ORCA or Gaussian09 and sometimes reading wavefunctions from other software like Tonto using the input script provided in listing 3.1.

A.9 Olex2

The development version of Olex2 was used, which was obtained directly from the developers and modified in the course of the development of NoSpherA2. [31] olex2.refine was used of these versions. [50]

A.10 ORCA

ORCA version 4.1.2 and 4.2.1 were used in the course of this thesis. [54, 55] It was used for calculation of wavefunctions in single point SCF procedures and for geometry optimizations in gas phase or continuous solvation models, as well as the QM engine for QM/MM procedures.

A.11 Tonto

Tonto, of the latest versions on the SF-Printout branch or master and release branches were used during this thesis. [153]

A.12 VMD

VMD versions 1.9.3 and 1.9.4 were used for visualizations, setup of MD and QM/MM calculations and analysis of the FEP calculations. [56]

Appendix B

Enabling Quantum Crystallography

B.1 Supporting Information for *Publication:* Generalizing Non-Spherical Structure Refinement: Hirshfeld Atom Refinement in NoSpherA2

All refinements in this SM, as well as all figures, tables and text were produced by me. The original datasets were retrieved from the publication "*Fast and Accurate Quantum Crystallography: From Small to Large, from Light to Heavy*" and changed according to the mentioned procedure and re-refined. All data statistics were performed by me.

Supporting Information for
Accurate Crystal Structures and Chemical Properties from NoSpherA2

Florian Kleemiss, Oleg V. Dolomanov, Michael Bodensteiner, Norbert Peyerimhoff, Laura Midgley, Luc J. Bourhis, Alessandro Genoni, Lorraine A. Malaspina, Dylan Jayatilaka, John L. Spencer, Fraser White, Bernhard Grundkötter-Stock, Simon Steinhauer, Dieter Lentz, Horst Puschmann*, Simon Grabowsky*

Table S1: Crystallographic, measurement and refinement details of the compounds measured for this study

	New Measurements			*Re-Refinements*	
Structure	$OsH_6(PC_{12}H_{19})_2$	CaF_2	$B_6H_6\ (NH_4)_2$	$C_{10}H_{10}N_4F_2$	$(C_6H_6O_2)(CO_2)_{0.854}$
Space group	P 2_1/n	Fm-3m	Fm-3m	P 2_1/n	R-3
a /Å	10.8918(1)	5.45095(5)	9.10295(12)	6.9196(1)	16.1737(2)
b /Å	13.7619(1)	5.45095(5)	9.10295(12)	14.5749(2)	16.1737(2)
c /Å	17.0714(2)	5.45095(5)	9.10295(12)	9.7248(1)	5.7050(1)
β /°	98.563(1)	90	90	90.637(1)	90
V /Å^3	2530.34(4)	161.963(3)	754.304(17)	980.71(2)	1292.42(3)
T /K	120	100	100	120	100
Resolution /Å	0.58	0.40	0.40	0.70	0.45
Wavelength /Å	0.71073	0.56087	0.71073	0.71073	0.71073
R_{int}	0.0582	0.0673	0.0243	0.0519	--
Avg. redundancy	19.05	55.20	11.91	4.77	1
Completeness	1.00	1.00	1.00	1.00	0.997
Average I/σ	40.3	85.1	85.5	35.1	39.4
# of refln. measured	196209	5299	4336	13911	3389
# of unique refln.	13109	96	364	2975	3388
Criterium for observed refln.	$F_o^2 > 2\sigma(F_o^2)$				
# of observed refln.	11278	96	364	2456	2937
Weighting scheme	0.0005/0.2531	0.0175/ 0.0607	0.0110/0.0085	0.0177/0.0	0.0295/0.0367
Level of theory	DKH2-PBE/x2c-TZVPP	PBE/def2-TZVPP			
QM-software	ORCA				
Partitioning code/ accuracy	NoSpherA2 / High				
Wfn options	SlowConv/NormalSCF	NormalConv/NormalSCF			
Other options:	Relativistics, No AFIX, DISP	DISP	No AFIX, H Aniso, DISP	DISP	No AFIX, H Aniso, DISP
Final R_1	0.0119	0.0114	0.0095	0.0214	0.0253
Final wR_2	0.0185	0.0279	0.0234	0.0483	0.0655
Max residual density /eÅ^{-3}	1.2148	0.6860	0.1093	0.1588	0.5303
Min residual density /eÅ^{-3}	-0.8989	-0.4146	-0.0662	-0.2130	-0.2211
CCDC deposition number	2034388	2034386	2034385	2034387	2034389

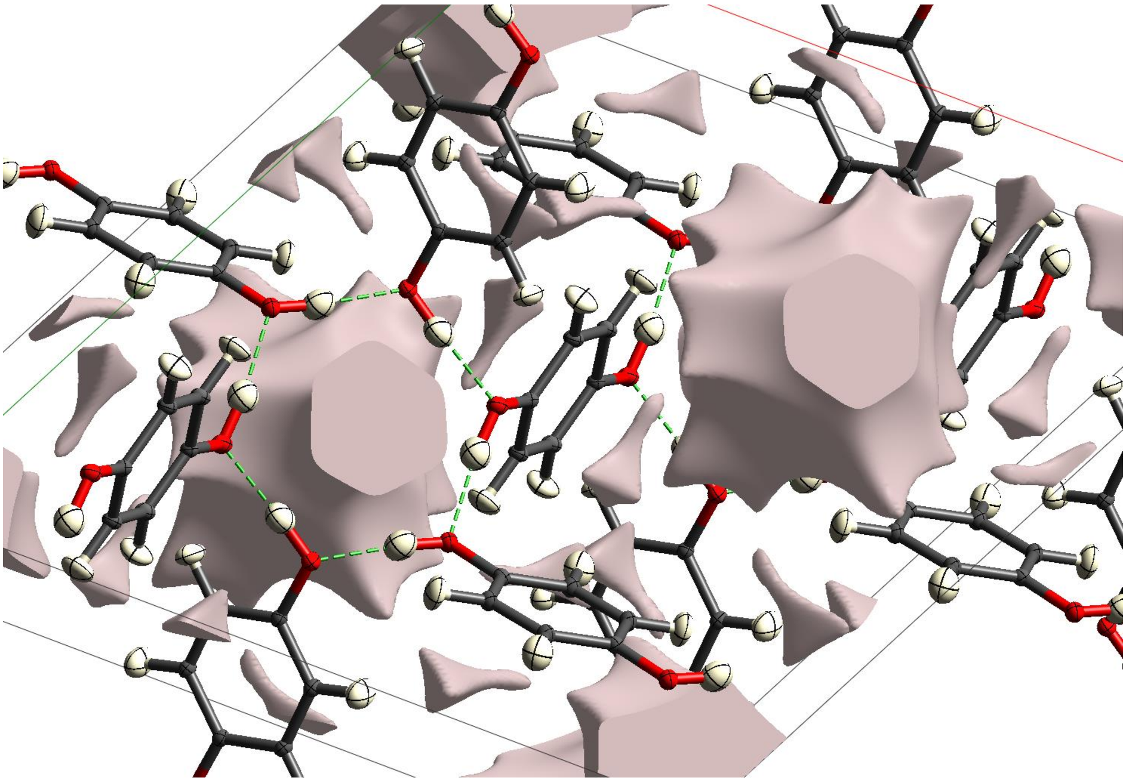


Figure S1: Visualization of voids based on 0.002 a.u. ED isosurfaces of the unit cell of HQ-CO2 if the CO_2 molecules are artificially removed from the structure.

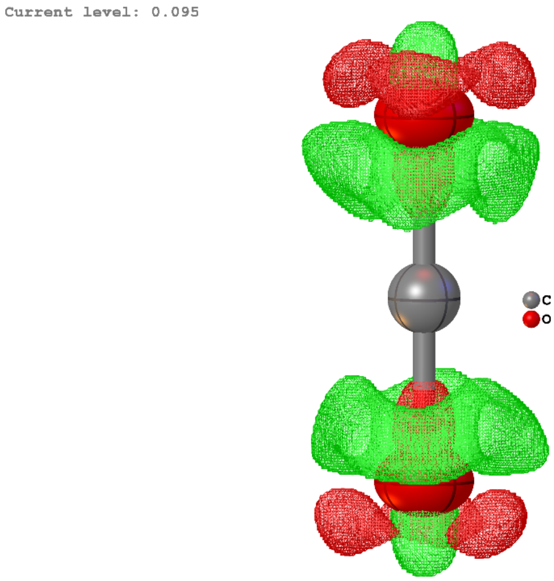


Figure S2: Residual density around the CO_2 molecule inside HQ signaling anharmonic motions. Iso-value = 0.10 e$Å^{-3}$ (green = positive, red = negative).

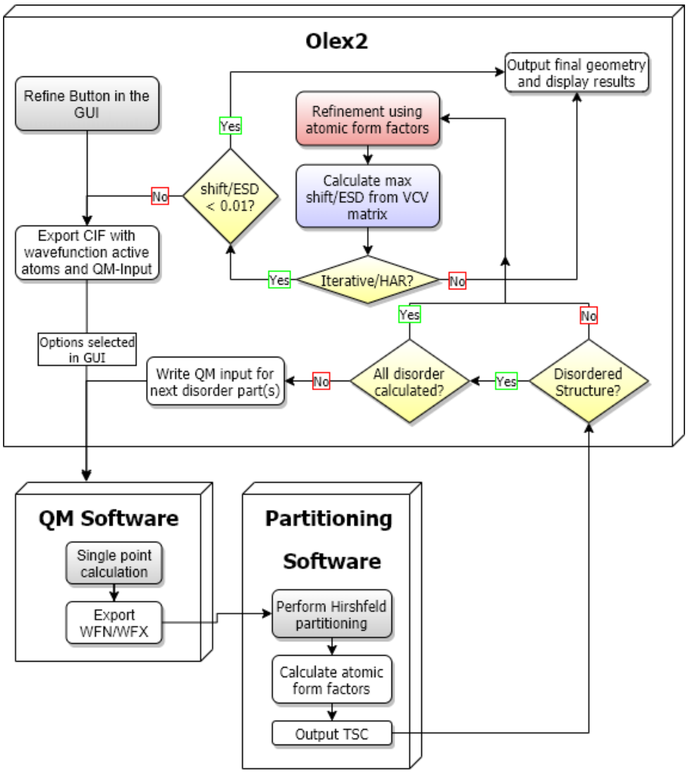
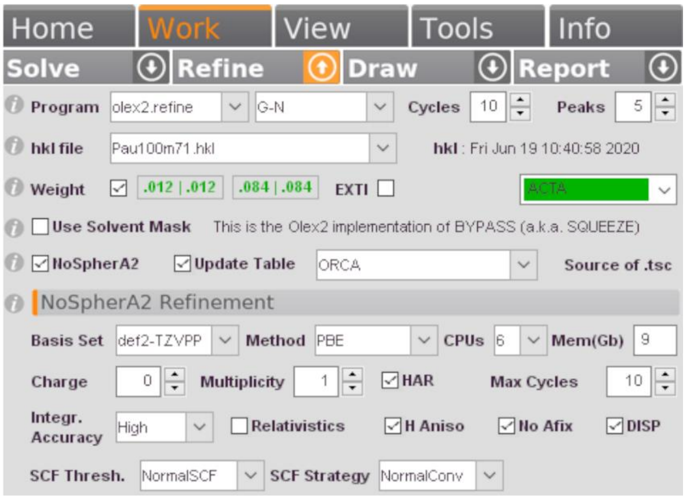


Figure S3: Detailed flowchart of the process behind HAR controlled by *NoSpherA2*.



Program olex2.refine G-N Cycles 10 Peaks 5

hkl file Pau100m71.hkl hkl : Fri Jun 19 10:40:58 2020

Weight ☑ .012 | .012 .084 | .084 EXTI ☐ ACTA

☐ Use Solvent Mask This is the Olex2 implementation of BYPASS (a.k.a. SQUEEZE)

☑ NoSpherA2 ☑ Update Table ORCA Source of .tsc

NoSpherA2 Refinement

Basis Set def2-TZVPP Method PBE CPUs 6 Mem(Gb) 9

Charge 0 Multiplicity 1 ☑ HAR Max Cycles 10

Integr. Accuracy High ☐ Relativistics ☑ H Aniso ☑ No Afix ☑ DISP

SCF Thresh. NormalSCF SCF Strategy NormalConv

Figure S4: Graphical user interface of *NoSpherA2* inside *Olex2*.

Theoretical framework for non-spherical structure refinement:

This section is an adaptation of reference 65.

Introduction

We have implemented a procedure that allows the use of non-spherical atomic form factors in crystallographic refinement. It is agnostic to the method employed to compute those form factors, as the refinement engine *olex2.refine*[S1] will use tabulated atomic form factors if the file specified below is present. The refinement will proceed as expected from an accomplished crystallographic refinement program, including the ability to make use of constraints, restraints, disorder and other specific tools like twin refinement and solvent masking. This procedure is available from all versions of *Olex2-1.3* and coupled to *NoSpherA2* as described in the main text.[S2]

Crystallographic refinement typically treats atoms as isolated, stand-alone entities with a spherically symmetrical electron charge distribution. A non-spherical treatment arises naturally from the fact that the electron density distribution of an atom is influenced by its environment. Historically, Stewart derived non-spherical form factors for bonded hydrogen atoms and commented: “By necessity, if not by choice, crystallographers have treated bonded atoms as point nuclei with a spherically symmetrical distribution of electron charge”.[S3]

The non-spherical form factors of the individual atoms - and their tabulation in the required format - can be obtained by various means. A possible starting point could be a molecular quantum mechanical wave-function calculation, followed by its transformation into electron densities and then the partitioning into atomic contributions. Alternatively, the electron densities themselves can be approximated using data-based contributing fragments.

Refinements based on non-spherical form factors are now possible in *olex2.refine*, and we wish to provide a rigorous mathematical justification for refinements using non-spherical form factors in this way. We would like to clarify that those form factors are not refined by *olex2.refine*. Only the usual parameters (positions, ADPs, occupancies, etc) are refined. These form factors can (and must) be externally recomputed after each refinement cycles, so that the form factor of each atom keeps matching the chemical environment as it changes during refinement.

Tabulated Atomic Form factors

Olex2 and *NoSpherA2* expect a file called *[name].tsc* (matching the *.hkl* file name) containing the following information in order to use the external atomic form factors:

The header of the *[name].tsc* file is free-format, as long as it contains the space-separated list of atom names in the `SCATTERERS:' line and finishes with `DATA:'. Any identifier must be followed by a colon. The identifiers may start with a space.

TITLE: *optional title of the structure*
SYMM: `expanded' *or list of symmetries*[1]
AD: TRUE or FALSE (anomalous dispersion)
SCATTERERS: *space-separated list of all atoms*
[ANYTHING]: *colon must be present*
DATA: (denotes the end of the header)

h	k	l	A₁	A₂	…	Aₙ
h_1	k_2	l_1	$f_1(h_1,k_1,l_1)$	$f_2(h_1,k_1,l_1)$	…	$f_n(h_1,k_1,l_1)$
h_2	k_2	l_2	$f_1(h_2,k_2,l_2)$	$f_2(h_2,k_2,l_2)$	…	$f_n(h_2,k_2,l_2)$
⋮	⋮	⋮	⋮	⋮	⋮	⋮
h_n	k_n	l_n	$f_1(h_n,k_n,l_n)$	$f_2(h_n,k_n,l_n)$	…	$f_n(h_n,k_n,l_n)$

As will be shown in the theory section, $f_j(h_i,k_i,l_i)$ is the form factor (Fourier transform of the electron density) of the atom A_j calculated in a coordinate system obtained by translating the origin of the crystallographic axes to the centre of atom A_j, at h_i, k_i, l_i. Index $j \in (1,\cdots,N)$ should run over all unique atoms of the asymmetric unit, and $i \in (1,\cdots,m)$ should run over at least all reflections defined in the *.hkl* file and any equivalents under symmetry.

The complex values $f_j(h_i,k_i,l_i)$ must be written as "Re,Im" - their real component followed by a comma followed by the imaginary component, with no spaces.

The format and information specified in the *.tsc* files was motivated by the mathematical derivations presented in the next section.

Theory

We will explain the mathematics behind the use of non-spherical form factors and how *olex2.refine* has adapted to enable their use. We keep the notation close to the one used in Reference [S1]. We will first discuss the standard case and then briefly discuss the modifications needed for twinning.

Monocrystals

Our mathematical arguments focus on the necessary modifications concerning the treatment of the calculated structure factor, which we denote by $F(\mathbf{h},\mathbf{y}(\mathbf{x}))$. Here $\mathbf{h}$ (a row vector) is a triplet of Miller indices, $\mathbf{y}$ comprises the crystallographic parameters (atomic positions and atomic displacement parameters (ADPs) and chemical occupancies) and the refinement is carried out with respect to potentially reduced parameters denoted by $\mathbf{x}$. The dependency of $\mathbf{y}$ on $\mathbf{x}$ is known analytically, and we emphasise this dependency by writing $\mathbf{y}$ as $\mathbf{y}(\mathbf{x})$, which does therefore embody all constraints.

The structure factor is the sum of the individual contributions of all atoms, partitioned into symmetry equivalent atoms to representatives A_j, $j = 1,2,\dots,N$, in the asymmetric unit

$$F(\mathbf{h},\mathbf{y}(\mathbf{x})) = \sum_{j=1}^{N} \underbrace{\left(\sum_{(R|\mathbf{t})\in S} s_j f_j^{(R|\mathbf{t})}(\mathbf{h},\mathbf{y}(\mathbf{x})) \right)}_{\text{atoms equivalent by symmetry to } A_j} \tag{1}$$

---

[1] In either case, all symmetry equivalent Miller indices must be present in the DATA section. If a list of symmetry operators, expressed as rotation matrices (e.g.: 1 0 0 0 1 0 0 0 1;-1 0 0 0 1 0 0 0 -1) is provided, then the Miller indices must be ordered into corresponding blocks - and each block must have symmetry equivalent indices in the same position in each block and generated by the corresponding matrices. This allows for more efficient calculations during the refinement. Otherwise, if SYMM has the value 'expanded', the indices can be present in any order.

with R the rotational part and $\mathbf{t}$ the translational part of the symmetry operation $(R|\mathbf{t}) \in \boldsymbol{S}$, and $\boldsymbol{s_j}$ the chemical occupancy of the representative $\boldsymbol{A_j}$.
The representative atoms $\boldsymbol{A_j}$ lie at fractional locations $\mathbf{z}_j$ (a column vector) with atomic vibration tensor U_j (a 3×3 symmetric matrix).[2] This information is contained within the vector $\mathbf{y}(\mathbf{x})$. For the chosen representative atom A_j in the asymmetric unit, its individual contribution $f_j^{(\mathbb{1}|\mathbf{0})}$ is given by

$$\begin{aligned} f_j^{(\mathbf{1}|\mathbf{0})}(\mathbf{h},\mathbf{y}(\mathbf{x})) &= f_j(\mathbf{h},\mathbf{y}(\mathbf{x}))e^{-\mathbf{h}U_j\mathbf{h}^T}e^{i2\pi\mathbf{h}\mathbf{z}_j} \\ &= f_j(\mathbf{h},\mathbf{y}(\mathbf{x}))\, G_j\left(\mathbf{h},\mathbf{y}_j(\mathbf{x})\right). \end{aligned} \tag{2}$$

Here $\mathbf{y}_j(\mathbf{x})$ is the subset of parameters of the structure pertaining to the j-th atom (namely U_j and $\mathbf{z}_j$), whilst the terms $f_j(\mathbf{h},\mathbf{y}(\mathbf{x}))$ are derived from the complex values $f_j(h_i,k_i,l_i)$ given in the .tsc-file by various corrections like, for example, anomalous dispersion, extinction etc.
The form factor of the atom A_j is calculated in a coordinate system obtained by translating the origin of the crystallographic axes to the centre of atom A_j, with no change in orientation. The form factor $f_j(\mathbf{h},\mathbf{y}(\mathbf{x}))$ is then the Fourier transform of the electron density ρ_j of A_j. In contrast to the case of spherical form factors, this electron density can now depend on the whole structure whose information is given in $\mathbf{y}(\mathbf{x})$, as non-spherical form factors take the dependence of the electron density of the surrounding atomic environment into account.
The relation between $f_j^{(R|\mathbf{t})}$ for a general $(R|\mathbf{t}) \in \boldsymbol{S}$ and f_j is then

$$\begin{aligned} f_j^{(R|\mathbf{t})}(\mathbf{h},\mathbf{y}(\mathbf{x})) &= f_j(\mathbf{h}R,\mathbf{y}(\mathbf{x}))e^{-\mathbf{h}RU_jR^T\mathbf{h}^T}e^{i2\pi\mathbf{h}R\mathbf{z}_j}e^{i2\pi\mathbf{h}\mathbf{t}} \\ &= f_j(\mathbf{h}R,\mathbf{y}(\mathbf{x}))G_j\left(\mathbf{h}R,\mathrm{y_j}(\mathbf{x})\right)\mathrm{e}^{\mathrm{i}2\pi\mathbf{h}\mathbf{t}} \\ &= f_j(\mathbf{h}R,\mathbf{y}(\mathbf{x}))\, G_j^{(R|\mathbf{t})}\left(\mathbf{h},\mathbf{y}_j(\mathbf{x})\right). \end{aligned} \tag{3}$$

Note that in the case of spherical form factors, the functions f_j do not depend on the structure information $\mathbf{y}(\mathbf{x})$ and, additionally, we have $f_j(\mathbf{h}R) = f_j(\mathbf{h})$ since $f_j(\mathbf{h})$ does then not depend on the direction of $\mathbf{h}$ but only on $\mathbf{h}M^*\mathbf{h}^T$, where M^* is the reciprocal metric matrix.
The least square minimization in the refinement procedure requires derivatives of the structure factor with respect to the components of $\mathbf{x} = (x_1, \ldots, x_n)$. Since the structure factor is the above sum (1), we only need to consider the derivatives of the individual terms $f_j^{(R|\mathbf{t})}(\mathbf{h},\mathbf{y}(\mathbf{x}))$.
Using the product rule, we have for the derivative (by dropping the arguments $\mathbf{y}(\mathbf{x})$ for ease of reading)

$$\frac{\partial f_j^{(R|\mathbf{t})}}{\partial x_k}(\mathbf{h}) = \frac{\partial f_j}{\partial x_k}(\mathbf{h}R)G_j^{(R|\mathbf{t})}(\mathbf{h}) + f_j(\mathbf{h}R)\,\frac{\partial G_j^{(R|\mathbf{t})}}{\partial x_k}(\mathbf{h}). \tag{4}$$

The partial differentiation with respect to x components will make partial derivatives of $\boldsymbol{y}(\boldsymbol{x})$ appear through Leibniz rule, as the time-honed implementation of constraints commands. It is important to note that the differential $\frac{\partial G_j^{(R|\boldsymbol{t})}}{\partial x_k}$ in the second term involves only exponentials which can be treated identically to the spherical case.

---

[2] The U_j here agree with $2\pi^2 U_j^*$ as introduced in reference S2.

The differential $\frac{\partial f_j}{\partial x_k}(\mathbf{h}R)$ in the first term on the right hand side of (4) is more difficult to treat due to the complexity of the involved derivations. We assume that the effect of this term for the least square minimisation procedure is relatively minor and thus take it as zero. The errors introduced via this and other assumptions will escalate if the structure changes without frequent updating of the non-spherical form factors. The validity of this assumption is expected to assert itself through the experimental exploration of this refinement technique in the field.

Given these considerations, the contributions of all symmetry equivalent atoms in both the structure factor and its derivatives require only the table $f_j(\mathbf{h})$ of the current structure information $\mathbf{y}(\mathbf{x})$ for each single representative A_j for each step of the refinement procedure. The experimental input to the refinement is a list of $\mathbf{h}$, $F_o^2(\mathbf{h})$ and $\sigma_o(\mathbf{h})$, where the last two items are respectively the measured intensities (scaled and with absorption corrections) and its estimated standard uncertainty (the *.hkl* file). Refinement is then a non-linear least squares (NLS) fit of $\left|F(\mathbf{h},\mathbf{y}(\mathbf{x}))\right|^2$ to $F_o^2(\mathbf{h}_j)$, for all $\mathbf{h}$. Precisely, the objective function to minimise is

$$\boldsymbol{x} \mapsto \sum_{r=1}^{m} w(\mathbf{h}_r)\left(\left|F(\mathbf{h}_r,\mathbf{y}(\mathbf{x}))\right|^2 - F_o^2(\mathbf{h}_r)\right)^2, \tag{5}$$

for m measured reflections, where $w(\mathbf{h}_r)$ are suitable weights. For more detail, see Section 2 of Reference [S1].

The set of Miller indices $\mathbf{h}$ required for the tabulated non-spherical form factors (the *.tsc* file) should correspond to the set of measured Bragg reflections and their symmetry equivalents.

Twinning

In the case of twinning, one needs to ensure that the set of Miller indices to be considered contains the measured Bragg peaks for all twin components and their symmetry equivalents. The modification required to the Least Squares (5) is to replace each term $\left|F(\boldsymbol{h}_r,\boldsymbol{y}(\boldsymbol{x}))\right|^2$, $r = 1,\dots,m$, by a combination over the contributing twin components indexed by l, namely

$$\sum_{l=1}^{d_r} \alpha_l \left|F\left(\boldsymbol{h}_{r,l},\boldsymbol{y}(\boldsymbol{x})\right)\right|^2 \tag{6}$$

where $d_r \leq d$ is the number of contributing components, α_l the fraction of the crystal volume occupied by the l-th contributing twin domain to the reflection h_r, and $h_{r,l}$ is the corresponding Miller index of this twin component contributing to this reflection. Note by equations (1) and (3), the calculation of $F\left(\mathbf{h}_{r,l},\mathbf{y}(\mathbf{x})\right)$ requires information of the non-spherical form factors for $\mathbf{h}_{r,l}$ and all its symmetry equivalents.

For more information on the general twinning procedure, see Section 5 of Reference [S1].

Therefore, all types of twinning can be handled by providing as input, for each reflection $\mathbf{h}_r$, the corresponding Miller indices $\mathbf{h}_{r,l}$ of the contributing components $l \in \{1,\dots,d\}$, and the matching $F_o^2(\mathbf{h}_r)$ and $\sigma_o(\mathbf{h}_r)$.

For the computation of structure factors using non-spherical form factors, only those Miller indices $\boldsymbol{h}_{r,l}$ and their symmetry equivalents are necessary. It is of course well known that in the case of (pseudo-)merohedral twinning, the Miller indices for a given r and varying l are related to each other by a twin law, but this is only a special case of the general scheme we have just described: the calculation of form factors does not need to be aware of this detail.

Summary

Let us finally cover the relevant differences to be considered when working with non-spherical form factors:

i. Form factors associated to atoms (with the origin at their center) are no longer real, but are usually complex-valued (as the electron densities are non-spherical).
ii. It is no longer the case that $f_j(\mathbf{h}R) = f_j(\mathbf{h})$ for rotations R associated to symmetry equivalent atoms in the unit cell.
iii. Due to the change in the shape of form factors under shifts, there appears an additional term in the derivative of $f_j^{(R|\mathbf{t})}$, the first term on the right hand side of (4). We assume that this is negligible for sufficiently small shifts.
iv. The provided form factors must cover a greater variety of Miller indices than would be needed in the spherical case (due to (ii)). In other words, form factors must be provided for all Miller indices $\mathbf{h}$ with recorded reflections and all symmetry equivalents $R\mathbf{h}$, for $(R, \mathbf{t})$ appearing in $\boldsymbol{S}$ (see (1)). This is also relevant for dealing with twin laws.

Validation Part:

Measurement details

The results presented here involve the comparison of the results obtained through HAR using *Tonto* and HAR using *NoSpherA2* for the following X-ray datasets:

- Glycyl-*L*-alanine, at 12K, λ= 0.5259 (2) Å, data set taken from ref. S5
- Glycyl-*L*-alanine, at 50K, λ= 0.5259 (2) Å, data set taken from ref. S5
- Glycyl-*L*-alanine, at 100K, λ= 0.5259 (2) Å, data set taken from ref. S5
- Glycyl-*L*-alanine, at 150K, λ= 0.5259 (2) Å, data set taken from ref. S5
- Glycyl-*L*-alanine, at 295K, λ= 0.5259 (2) Å, data set taken from ref. S5
- *L*-Alanine, at 23K, λ= 0.71073 Å, data set taken from ref. S6
- *L*-Alanine, at 100K, λ= 0.71073 Å, results taken from ref. S7
- *L*-Alanine, at 150K, λ= 0.71073 Å, results taken from ref. S7

Neutron-diffraction data sets are available for all structures except for L-alanine and glycyl-*L*-alanine at 100K.

- Glycyl-*L*-alanine, at 12K, λ= 0.8313(2) Å, data set taken from ref. S8
- Glycyl-*L*-alanine, at 50K, λ= 0.8313(2) Å, data set taken from ref. S8
- Glycyl-*L*-alanine, at 150K, λ= 0.8313(2) Å, data set taken from ref. S8
- Glycyl-*L*-alanine, at 295K, λ= 0.8313(2) Å, data set taken from ref. S8
- *L*-Alanine, at 23K, λ= 0.750 to 1.500 Å (Laue), results taken from ref. S7
- *L*-Alanine, at 150K, λ= 0.750 to 1.500 Å (Laue), results taken from ref. S7

Refinement results

The R factor was calculated on the basis of manually edited *.hkl* files containing an unusual cut-off criterion of 2.1 I/σ(I) to ensure neither *olex2.refine* nor *Tonto* would disregard any reflections during refinement, statistics preparation or residual density calculation. R-factor and Goodness of Fit (S) were taken from CIFs, but the residual density was calculated using *Olex2* using *.fcf* files.

HARs of Gly-L-Ala (CCDC-2035148)

Table S2: Refinements against X-ray data for Gly-*L*-Ala at 12 K

	Tonto	**NoSpherA2**	**Olex2.refine IAM**
R*[F^2>2.1σ(F^2)], *S	0.016, 1.26	0.016, 1.29	0.024, 2.45
No. of reflections	2374	2374	2374
No. of parameters	181	181	131
H-atom treatment	All H-atom parameters refined (anisotropically)	All H-atom parameters refined (anisotropically)	H atoms refined isotropically
$\Delta\rho_{max}$, $\Delta\rho_{min}$ (eÅ^{-3})	0.151, -0.194	0.152, -0.194	0.289, -0.253

Table S3: Refinements against X-ray data for Gly-*L*-Ala at 50 K

	Tonto	**NoSpherA2**	**Olex2.refine IAM**
R*[F^2>2.1σ(F^2)], *S	0.018, 1.21	0.018, 2.24	0.025, 2.16
No. of reflections	2261	2261	2261
No. of parameters	181	181	131
H-atom treatment	All H-atom parameters refined (anisotropically)	All H-atom parameters refined (anisotropically)	H atoms refined isotropically
$\Delta\rho_{max}$, $\Delta\rho_{min}$ (eÅ^{-3})	0.137, -0.180	0.137, -0.166	0.259, -0.232

Table S4: Refinements against X-ray data for Gly-*L*-Ala at 100 K

	Tonto	**NoSpherA2**	**Olex2.refine IAM**
R*[F^2>2.1σ(F^2)], *S	0.018, 1.57	0.018, 1.59	0.026, 3.11
No. of reflections	2430	2430	2430
No. of parameters	181	181	131
H-atom treatment	All H-atom parameters refined (anisotropically)	All H-atom parameters refined (anisotropically)	H atoms refined isotropically
$\Delta\rho_{max}$, $\Delta\rho_{min}$ (eÅ^{-3})	0.136, -0.254	0.135, -0.254	0.267, -0.257

Table S5: Refinements against X-ray data for Gly-*L*-Ala at 150 K

	Tonto	**NoSpherA2**	**Olex2.refine IAM**
$R[F^2>2.1\sigma(F^2)]$, S	0.016, 1.54	0.016, 1.57	0.025, 3.13
No. of reflections	2391	2391	2391
No. of parameters	181	181	131
H-atom treatment	All H-atom parameters refined (anisotropically)	All H-atom parameters refined (anisotropically)	H atoms refined isotropically
$\Delta\rho_{max}$, $\Delta\rho_{min}$ (eÅ^{-3})	0.124, -0.196	0.124, -0.196	0.221, -0.227

Table S6: Refinements against X-ray data for Gly-*L*-Ala at 295 K

	Tonto	**NoSpherA2**	**Olex2.refine IAM**
$R[F^2>2.1\sigma(F^2)]$, S	0.019, 1.44	0.019, 1.47	0.030, 2.81
No. of reflections	2265	2265	2265
No. of parameters	181	181	131
H-atom treatment	All H-atom parameters refined (anisotropically)	All H-atom parameters refined (anisotropically)	H atoms refined isotropically
$\Delta\rho_{max}$, $\Delta\rho_{min}$ (eÅ^{-3})	0.132, -0.184	0.132, -0.184	0.191, -0.214

HARs of L-Alanine (CCDC-2035147)

Table S7: Refinements against X-ray data for *L*-Ala at 23 K

	Tonto	**NoSpherA2**	**Olex2.refine IAM**
$R[F^2>2.1\sigma(F^2)]$, S	0.019, 1.18	0.019, 1.18	0.028, 2.31
No. of reflections	2387	2387	2387
No. of parameters	118	118	83
H-atom treatment	All H-atom parameters refined (anisotropically)	All H-atom parameters refined (anisotropically)	H atoms refined isotropically
$\Delta\rho_{max}$, $\Delta\rho_{min}$ (eÅ^{-3})	0.112, -0.127	0.113, -0.126	0.349, -0.209

Table S8: Refinements against X-ray data for *L*-Ala at 100 K

	Tonto	**NoSpherA2**	**Olex2.refine IAM**
R*[F^2>2.1σ(F^2)], *S	0.019, 2.08	0.019, 2.09	0.027, 4.16
No. of reflections	4549	4549	4549
No. of parameters	118	118	83
H-atom treatment	All H-atom parameters refined (anisotropically)	All H-atom parameters refined (anisotropically)	H atoms refined isotropically
$\Delta\rho_{max}$, $\Delta\rho_{min}$ (eÅ^{-3})	0.159, -0.330	0.159, -0.327	0.379, -0.255

Table S9: Refinements against X-ray data for *L*-Ala at 150 K

	HAR	**NoSpherA2**	**Olex2.refine IAM**
R(F)*[F^2>2.1σ(F^2)], *S	0.020, 1.76	0.020, 1.75	0.028, 2.62
No. of reflections	2584	2584	2584
No. of parameters	118	118	83
H-atom treatment	All H-atom parameters refined (anisotropically)	All H-atom parameters refined (anisotropically)	H atoms refined isotropically
$\Delta\rho_{max}$, $\Delta\rho_{min}$ (eÅ^{-3})	0.233, -0.197	0.232, -0.199	0.495, -0.238

Methodology for validation of NoSpherA2

To be comparable with previous studies [S6], the same level of theory was chosen in *Tonto* and *NoSpherA2*: RHF/6-311G(d,p) without cluster charges and dipoles.

Three types of agreement statistics were used in the following:

1) Mean ratios of values X in two models A and B, denoted by **$\langle X_A/X_B\rangle$**
2) Mean absolute differences of two quantities **$\langle|\Delta X_{A\text{-}B}|\rangle$**
3) Root mean square differences of **X** weighted by combined standard uncertainties **σ**.

The wRMSD is calculated as:

$$wRMSD = \langle\frac{(X_A - X_B)^2}{\sigma(X_A)^2 + \sigma(X_B)^2}\rangle^{\frac{1}{2}}$$

For a perfect agreement between two models one would expect the ratio of all parameters to become 1. The mean absolute difference should be 0 for perfect agreement, but this does not say anything about the spread of the values around this mean point. The wRMSD becomes 0 for identical numerical values (as $(X_A - X_B)^2$ becomes 0), and close to 1 for statistical agreement.

Model names (compare to Table 2 in the main document):

ID	Type	Program	QM	Partitioning	Weighting	Model name for the SI
i	IAM	olex2.refine	--	--	$1/\sigma^2(F^2)$	IAM
ii	HAR	Tonto	Tonto	Tonto	$1/\sigma(F)$	Tonto
iii	HAR	NoSpherA2	Tonto	Tonto	$1/\sigma^2(F^2)$	NoSpherA2 - T
iv	HAR	NoSpherA2	ORCA	Tonto	$1/\sigma^2(F^2)$	NoSpherA2 - O
v	HAR	NoSpherA2	ORCA	Tonto	Shelxl-type	NoSpherA2 - w
vi	HAR	NoSpherA2	ORCA	NoSpherA2	$1/\sigma^2(F^2)$	NoSpherA2 - Q

The results show almost perfect agreement between Tonto and NoSpherA2 refinements. The differences in bond lengths (Table S9) for *Gly-L-Ala* are at a maximum of 0.002 Å, while the smallest standard uncertainty of any of the bond distances is 0.005 Å. This difference can be due to numerical differences in the least-squares implementation, since HAR in *Tonto* is performed on F, while *olex2.refine* uses F^2. In the case of *L-Ala*, the maximum bond distance difference is 0.008 Å, while standard uncertainties for HAR and NoSpherA2 are 0.009 Å and 0.011 Å, respectively.
For hydrogen atom ADPs in *Gly-L-Ala*, the biggest difference between Tonto and NoSpherA2 refinements in all temperatures is 0.0031 Å^2 in U^{11}, while the standard uncertainties are 0.0123 Å^2 and 0.0124 Å^2, respectively. For non-hydrogen atom ADPs, the biggest difference is 0.000068 Å^2 for U^{22}, while the standard uncertainties in both methods for this particular ADP are 0.00025 Å^2 and 0.00026 Å^2.

In *L*-Ala the biggest difference in any hydrogen atom ADP for any considered temperature is 0.0086 Å^2 for U^{23}, while the standard uncertainties are 0.0044 Å^2 and 0.0053 Å^2. Non hydrogen atoms have a maximum difference of 0.00046 Å^2 for U^{13} while the standard uncertainties are 0.00010 Å^2 and 0.00012 Å^2. All these maximum differences appeared at the highest temperature, as the absolute value is higher and therefore statistical and numerical differences would be bigger.

To visualize the validation results, Figures S5 – S10 were prepared by using the data in Tables S10-S21.

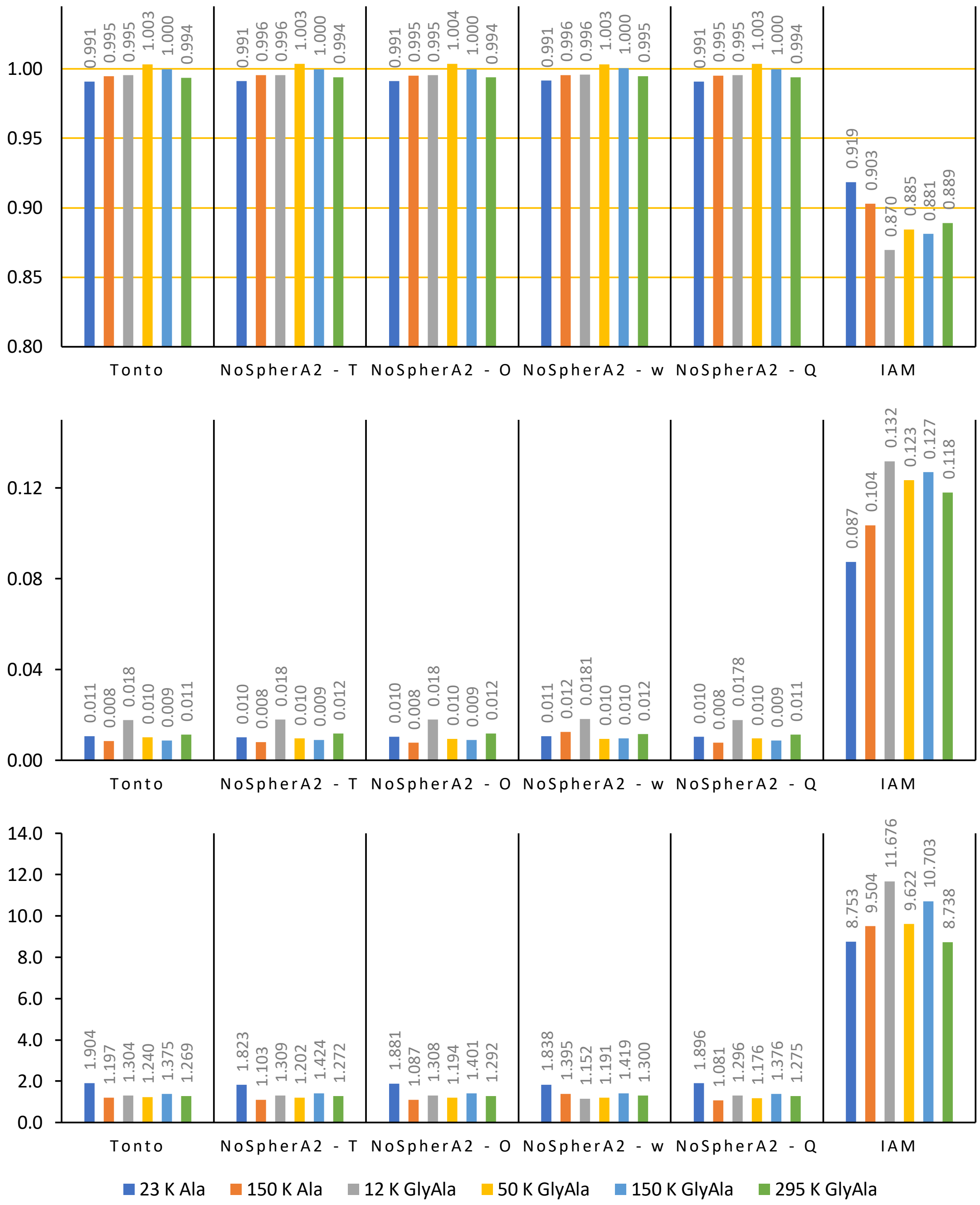


Figure S5: Plots of agreement in X-H bond lengths (Å) between different HAR models and neutron data. Top: ⟨$r_A/r_{Neutron}$⟩ (1 is best), middle: ⟨|$\Delta r_{\text{A-Neutron}}$|⟩ (0 is best), bottom: wRMSD (0 is best, 1 is statistical agreement)

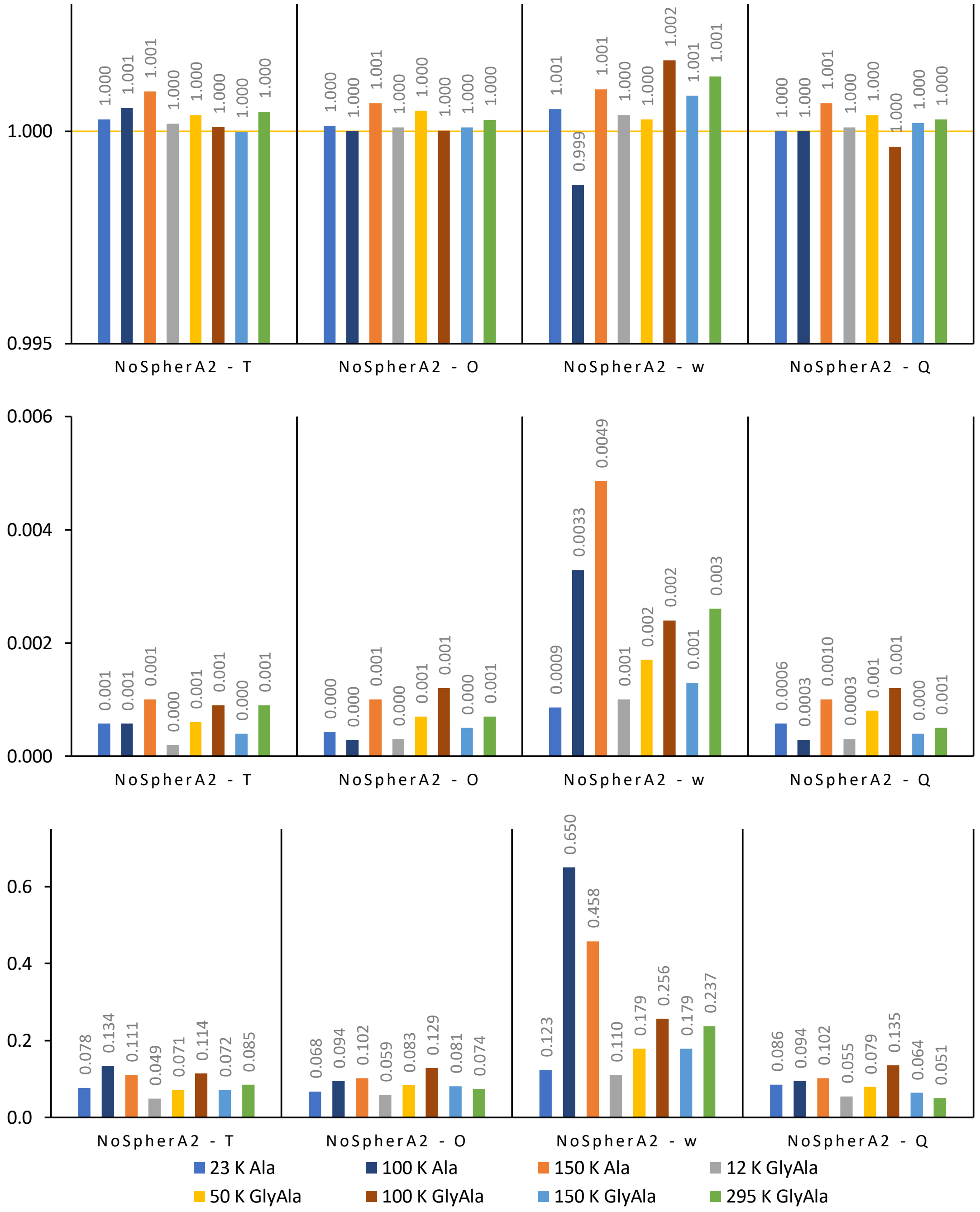


Figure S6: Plots of agreement in X-H bond lengths (Å) between different NoSpherA2 HAR models and HAR in Tonto. Top: $\langle r_A/r_{HAR}\rangle$ (1 is best), middle: $\langle|\Delta r_{\text{A-HAR}}|\rangle$ (0 is best), bottom: wRMSD (0 is best, 1 is statistical agreement)

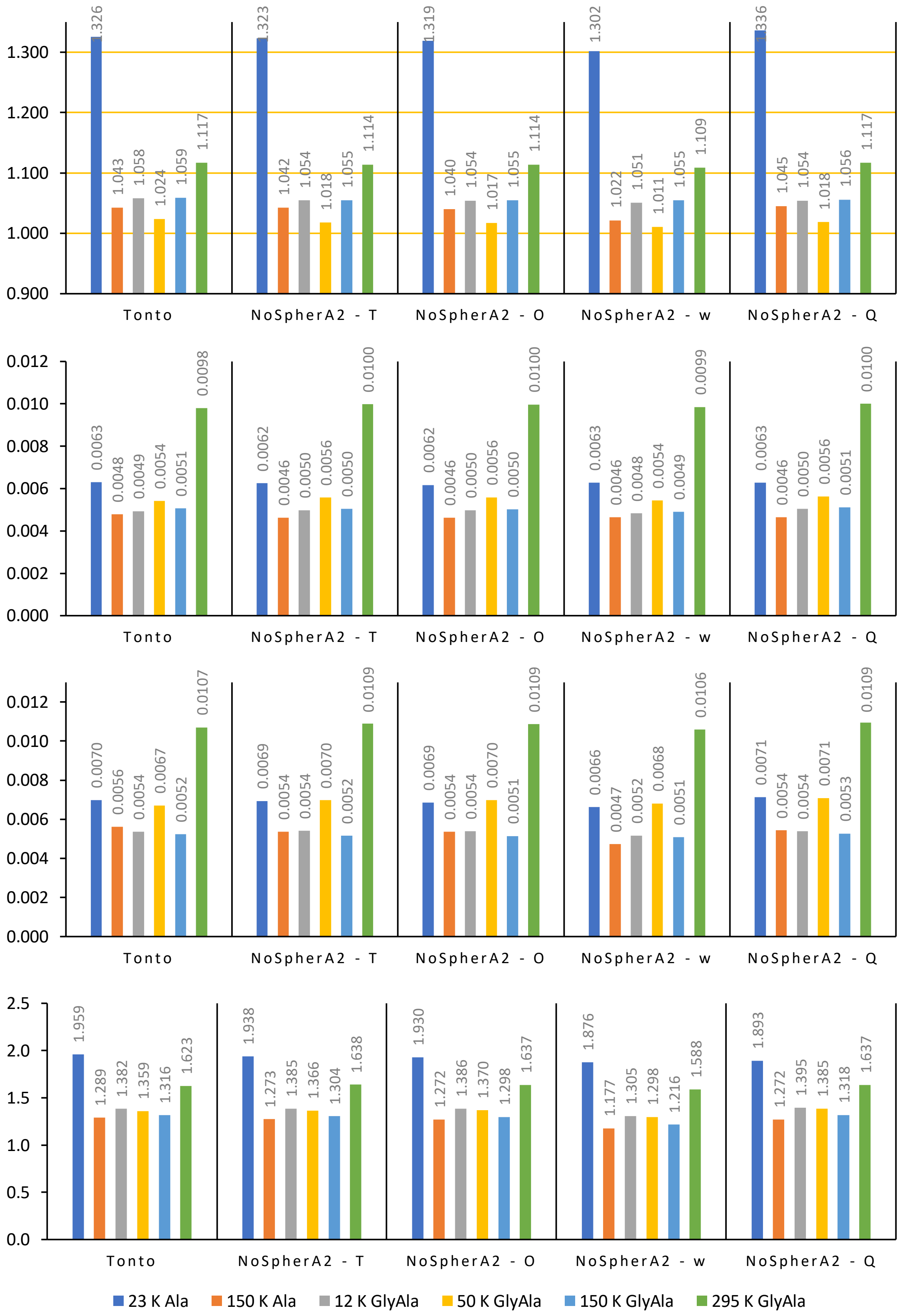


Figure S7: Plots of agreement in U^{ij} (Å^2) of hydrogen atoms between different HAR models and neutron data. From top to bottom: $\langle U^{ii}_{A}/U^{ii}_{Neutron}\rangle$ (1 is best), $\langle|\Delta U^{ij}_{\text{A-Neutron}}|\rangle$ (0 is best), $\langle|\Delta U^{ii}_{\text{A-Neutron}}|\rangle$ (0 is best), wRMSD (0 is best, 1 is statistical agreement)

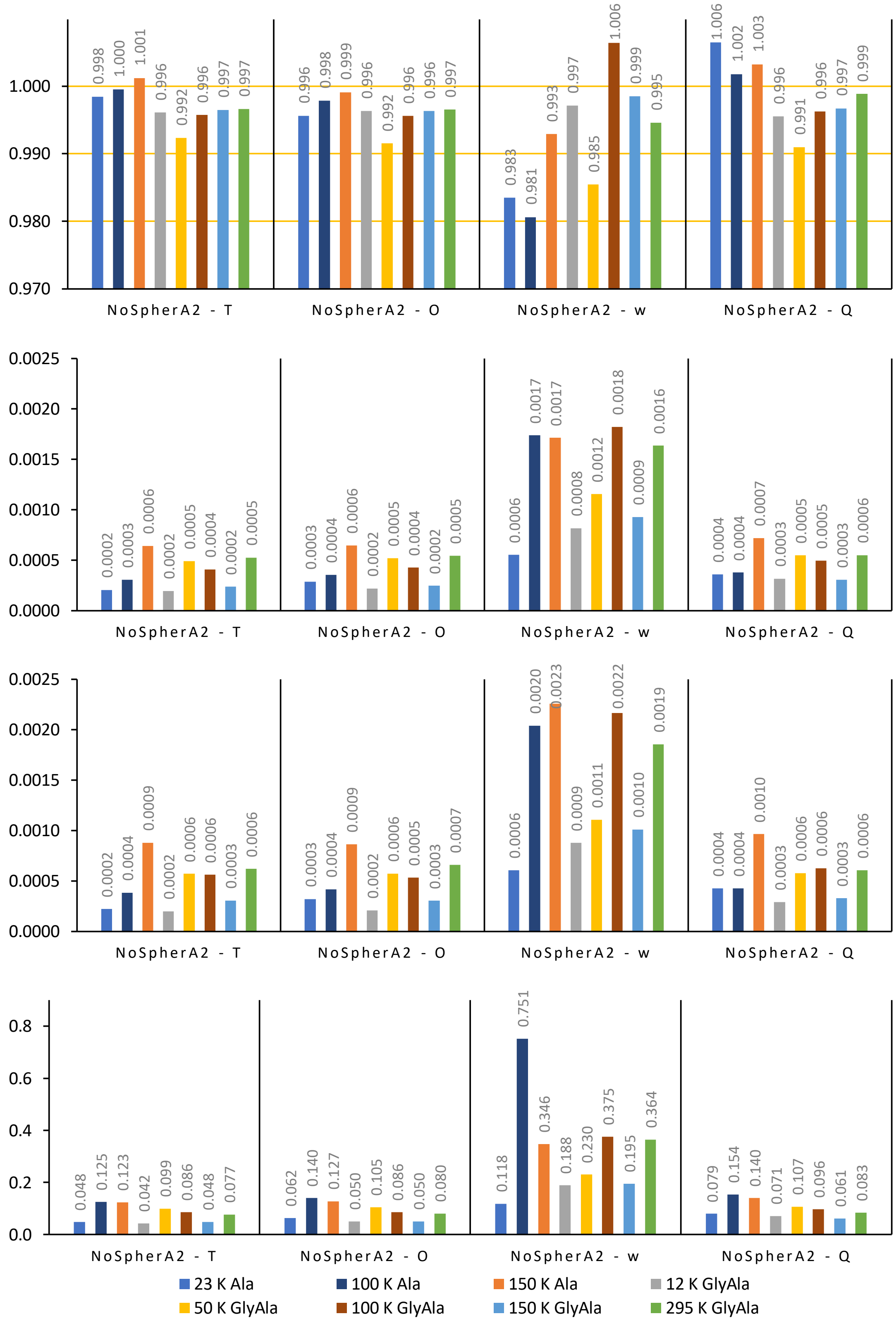


Figure S8: Plots of agreement in U^{ij} ($Å^2$) of hydrogen atoms between different NoSpherA2 HAR models and HAR in Tonto. From top to bottom: $\langle U^{ii}_{A}/U^{ii}_{HAR}\rangle$ (1 is best), $\langle|\Delta U^{ij}_{\text{A-HAR}}|\rangle$ (0 is best), $\langle|\Delta U^{ii}_{\text{A-HAR}}|\rangle$ (0 is best), wRMSD (0 is best, 1 is statistical agreement)

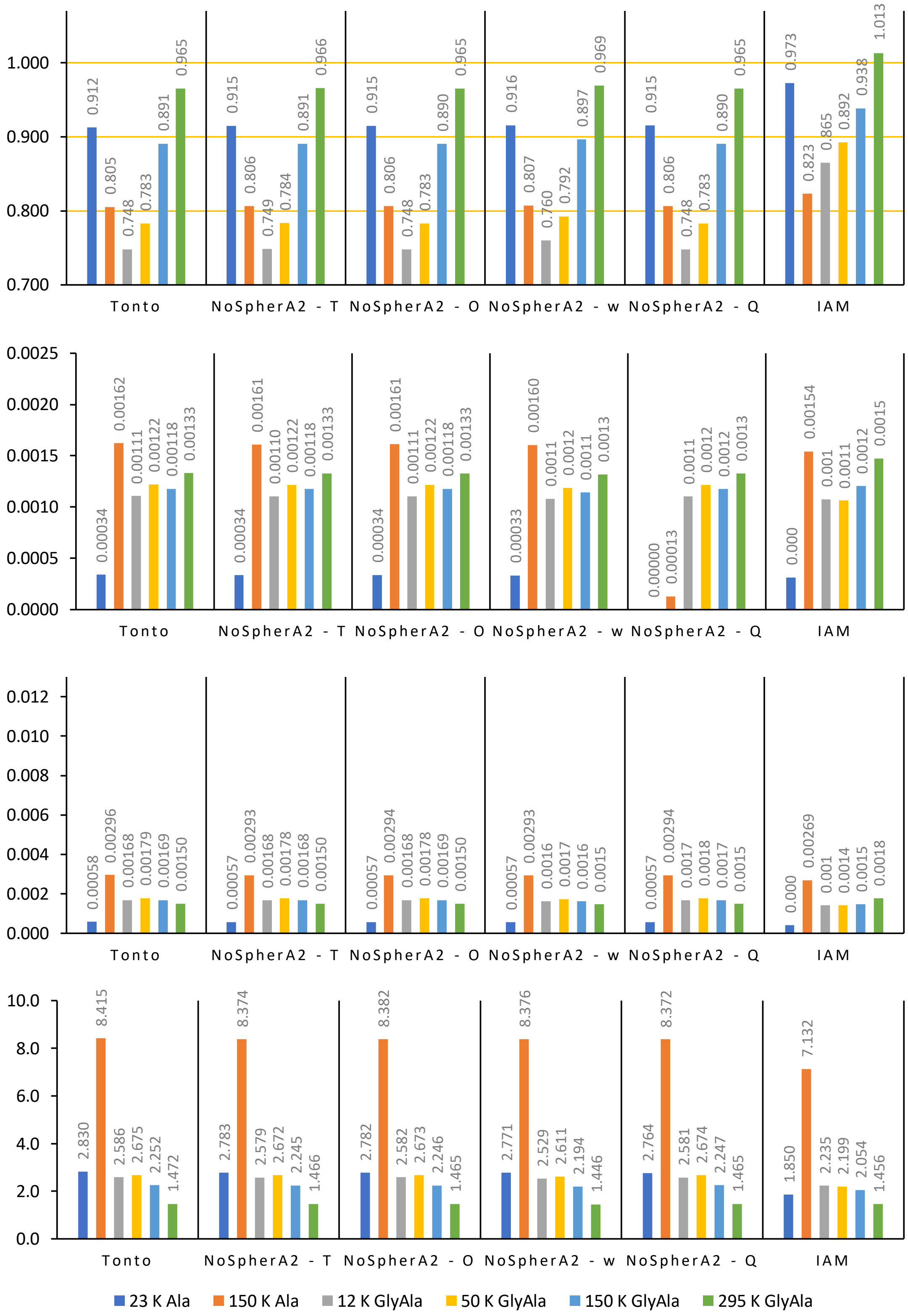


Figure S9: Plots of agreement in U^{ij} (Å^2) of non-hydrogen atoms between different HAR models and neutron data. From top to bottom: $\langle U^{ii}{}_{A}/U^{ii}{}_{Neutron}\rangle$ (1 is best), $\langle|\Delta U^{ij}{}_{\text{A-Neutron}}|\rangle$ (0 is best), $\langle|\Delta U^{ii}{}_{\text{A-Neutron}}|\rangle$ (0 is best), wRMSD (0 is best, 1 is statistical agreement)

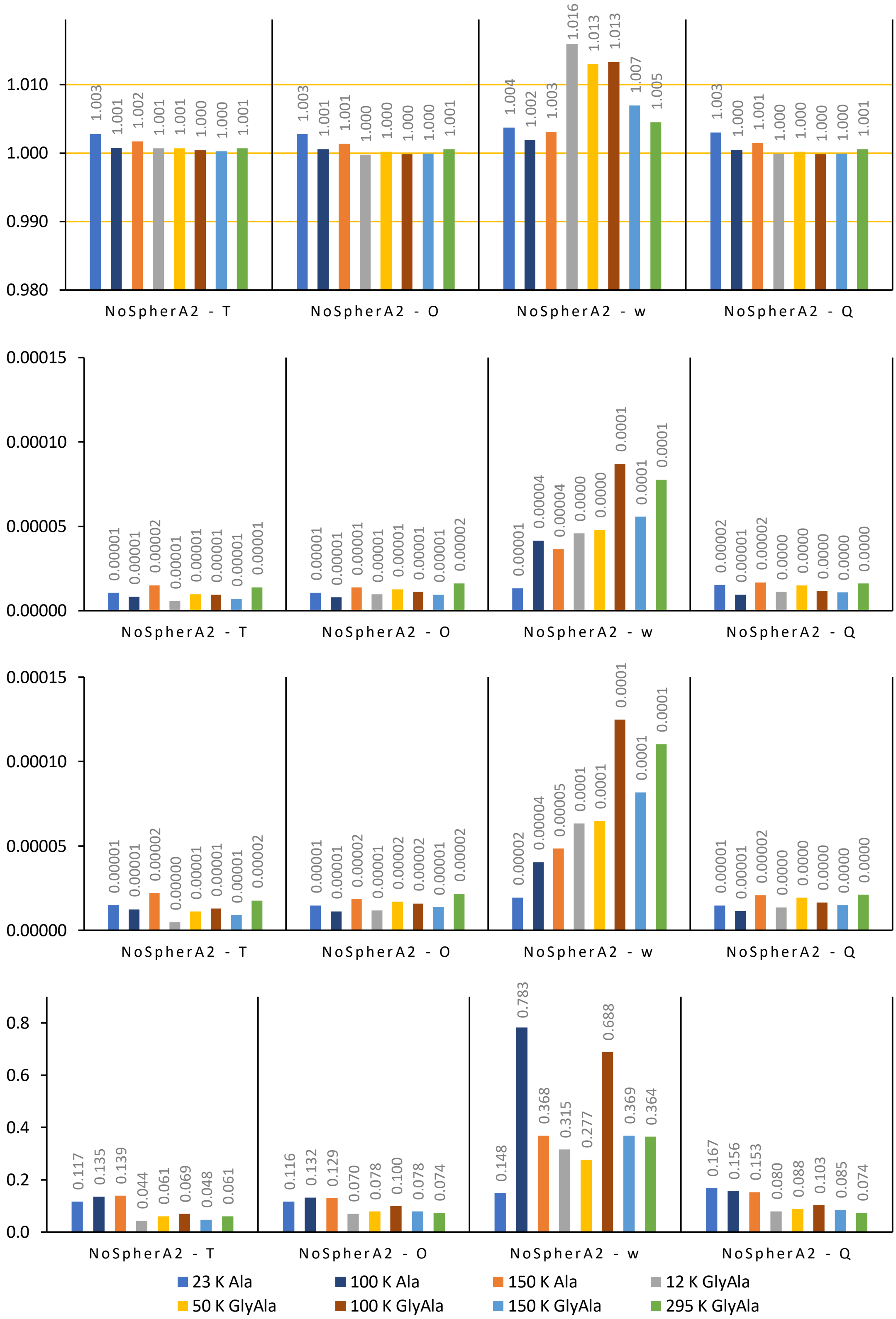


Figure S10: Plots of agreement in U^{ij} ($Å^2$) of non-hydrogen atoms between different NoSpherA2 HAR models and HAR in Tonto. From top to bottom: $\langle U^{ii}_{A}/U^{ii}_{HAR}\rangle$ (1 is best), $\langle|\Delta U^{ij}_{\text{A-HAR}}|\rangle$ (0 is best), $\langle|\Delta U^{ii}_{\text{A-HAR}}|\rangle$ (0 is best), wRMSD (0 is best, 1 is statistical agreement)

Table S10: Bond-length statistics for X-H distances, averaged among all bonds for Gly-*L*-Ala.

Temp.	**Method (A / B)**	$\langle r_A/r_B \rangle$	$\langle \lvert \Delta r_{\text{A-B}} \rvert \rangle$ / Å	**wRMSD**
12K	Tonto / Neutron	0.995	0.018	1.304
	NoSpherA2 – T / Neutron	0.996	0.018	1.309
	NoSpherA2 – O / Neutron	0.995	0.018	1.308
	IAM / Neutron	0.870	0.132	11.676
	Tonto / NoSpherA2 – T	1.000	0.000	0.049
	Tonto / NoSpherA2 – O	1.000	0.000	0.059
	NoSpherA2 – O / NoSpherA2 – T	1.000	0.000	0.032
50K	Tonto / Neutron	1.003	0.010	1.240
	NoSpherA2 – T / Neutron	1.003	0.010	1.202
	NoSpherA2 – O / Neutron	1.004	0.010	1.194
	IAM / Neutron	0.885	0.123	9.622
	Tonto / NoSpherA2 – T	1.000	0.001	0.071
	Tonto / NoSpherA2 – O	1.000	0.001	0.083
	NoSpherA2 – O / NoSpherA2 – T	1.000	0.000	0.025
100K	Tonto / Neutron	--	--	--
	NoSpherA2 – T / Neutron	--	--	--
	NoSpherA2 – O / Neutron	--	--	--
	IAM / Neutron	--	--	--
	Tonto / NoSpherA2 – T	1.000	0.001	0.114
	Tonto / NoSpherA2 – O	1.000	0.001	0.129
	NoSpherA2 – O / NoSpherA2 – T	1.000	0.000	0.060
150K	Tonto / Neutron	1.000	0.009	1.375
	NoSpherA2 – T / Neutron	1.000	0.009	1.424
	NoSpherA2 – O / Neutron	1.000	0.009	1.401
	IAM / Neutron	0.881	0.127	10.703
	Tonto / NoSpherA2 – T	1.000	0.000	0.072
	Tonto / NoSpherA2 – O	1.000	0.000	0.081
	NoSpherA2 – O / NoSpherA2 – T	1.000	0.000	0.037
295K	Tonto / Neutron	0.994	0.011	1.269
	NoSpherA2 – T / Neutron	0.994	0.012	1.272
	NoSpherA2 – O / Neutron	0.994	0.012	1.292
	IAM / Neutron	0.889	0.118	8.738
	Tonto / NoSpherA2 – T	1.000	0.001	0.085
	Tonto / NoSpherA2 – O	1.000	0.001	0.074
	NoSpherA2 – O / NoSpherA2 – T	1.000	0.000	0.042

Table S11: Bond-length statistics for X-H distances, averaged among all bonds for *L*-Ala.

Temp.	**Method (A / B)**	**$\langle r_A/r_B \rangle$**	**$\langle \lvert \Delta r_{\mathrm{A\text{-}B}} \rvert \rangle$ / Å**	**wRMSD**
23K	Tonto / Neutron	0.991	0.011	1.904
	NoSpherA2 – T / Neutron	0.991	0.010	1.823
	NoSpherA2 – O / Neutron	0.991	0.010	1.881
	IAM / Neutron	0.919	0.087	8.753
	Tonto / NoSpherA2 – T	1.000	0.001	0.078
	Tonto / NoSpherA2 – O	1.000	0.000	0.068
	NoSpherA2 – O / NoSpherA2 – T	1.000	0.000	0.066
100K	Tonto / Neutron	--	--	--
	NoSpherA2 – T / Neutron	--	--	--
	NoSpherA2 – O / Neutron	--	--	--
	IAM / Neutron	--	--	--
	Tonto / NoSpherA2 – T	1.001	0.001	0.134
	Tonto / NoSpherA2 – O	1.000	0.000	0.094
	NoSpherA2 – O / NoSpherA2 – T	0.999	0.001	0.134
150K	Tonto / Neutron	0.995	0.008	1.197
	NoSpherA2 – T / Neutron	0.996	0.008	1.103
	NoSpherA2 – O / Neutron	0.995	0.008	1.087
	IAM / Neutron	0.903	0.104	9.504
	Tonto / NoSpherA2 – T	1.001	0.001	0.111
	Tonto / NoSpherA2 – O	1.001	0.001	0.102
	NoSpherA2 – O / NoSpherA2 – T	1.000	0.000	0.042

Table S12: Statistical results for hydrogen atom ADPs for Gly-*L*-Ala, units of differences are Å^2.

Temp.	**Method (A/B)**	$< U^{ii}_{A} / U^{ii}_{B} >$	$< \| U^{ij}_{A} - U^{ij}_{B} \| >$	$< \| U^{ii}_{A} - U^{ii}_{B} \| >$	**wRMSD**
12 K	Tonto/Neutron	1.058	0.0049	0.0054	1.382
	NoSpherA2 - T/Neutron	1.054	0.0050	0.0054	1.385
	NoSpherA2 - O/Neutron	1.054	0.0050	0.0054	1.386
	NoSpherA2 - T/Tonto	0.996	0.0002	0.0002	0.042
	NoSpherA2 - O/Tonto	0.996	0.0002	0.0002	0.050
	NoSpherA2 - O/T	1.000	0.0001	0.0001	0.026
50 K	Tonto/Neutron	1.024	0.0054	0.0067	1.359
	NoSpherA2 - T/Neutron	1.018	0.0056	0.0070	1.366
	NoSpherA2 - O/Neutron	1.017	0.0056	0.0070	1.370
	NoSpherA2 - T/Tonto	0.992	0.0005	0.0006	0.099
	NoSpherA2 - O/Tonto	0.992	0.0005	0.0006	0.105
	NoSpherA2 - O/T	0.999	0.0001	0.0001	0.023
100 K	Tonto/Neutron	--	--	--	--
	NoSpherA2 - T/Neutron	--	--	--	--
	NoSpherA2 - O/Neutron	--	--	--	--
	NoSpherA2 - T/Tonto	0.996	0.0004	0.0006	0.086
	NoSpherA2 - O/Tonto	0.996	0.0004	0.0005	0.086
	NoSpherA2 - O/T	1.000	0.0001	0.0001	0.026
150 K	Tonto/Neutron	1.059	0.0051	0.0052	1.316
	NoSpherA2 - T/Neutron	1.055	0.0050	0.0052	1.304
	NoSpherA2 - O/Neutron	1.055	0.0050	0.0051	1.298
	NoSpherA2 - T/Tonto	0.997	0.0002	0.0003	0.048
	NoSpherA2 - O/Tonto	0.996	0.0002	0.0003	0.050
	NoSpherA2 - O/T	1.000	0.0001	0.0001	0.025
295 K	Tonto/Neutron	1.117	0.0098	0.0107	1.623
	NoSpherA2 - T/Neutron	1.114	0.0100	0.0109	1.638
	NoSpherA2 - O/Neutron	1.114	0.0100	0.0109	1.637
	NoSpherA2 - T/Tonto	0.997	0.0005	0.0006	0.077
	NoSpherA2 - O/Tonto	0.997	0.0005	0.0007	0.080
	NoSpherA2 - O/T	1.000	0.0001	0.0001	0.025

Table S13: Statistical results for non-hydrogen atom ADPs for Gly-*L*-Ala, units of differences are Å^2.

Temp.	**Method (A/B)**	**$< U^{ii}_A / U^{ii}_B >$**	**$< \| U^{ij}_A - U^{ij}_B \| >$**	**$< \| U^{ii}_A - U^{ii}_B \| >$**	**wRMSD**
12 K	Tonto/Neutron	0.748	0.00111	0.00168	2.586
	NoSpherA2 - T/Neutron	0.749	0.00110	0.00168	2.579
	NoSpherA2 - O/Neutron	0.748	0.00111	0.00168	2.582
	IAM/Neutron	0.865	0.00108	0.00141	2.235
	NoSpherA2 - T/Tonto	1.001	0.00001	0.00000	0.044
	NoSpherA2 - O/Tonto	1.000	0.00001	0.00001	0.070
	NoSpherA2 – O/T	0.999	0.00001	0.00001	0.060
50 K	Tonto/Neutron	0.783	0.00122	0.00179	2.675
	NoSpherA2 - T/Neutron	0.784	0.00122	0.00178	2.672
	NoSpherA2 - O/Neutron	0.783	0.00122	0.00178	2.673
	IAM/Neutron	0.892	0.00106	0.00142	2.199
	NoSpherA2 - T/Tonto	1.001	0.00001	0.00001	0.061
	NoSpherA2 - O/Tonto	1.000	0.00001	0.00002	0.078
	NoSpherA2 – O/T	0.999	0.00001	0.00001	0.052
100 K	Tonto/Neutron	-	-	-	-
	NoSpherA2 - T/Neutron	-	-	-	-
	NoSpherA2 - O/Neutron	-	-	-	-
	IAM/Neutron	-	-	-	-
	NoSpherA2 - T/Tonto	1.000	0.00001	0.00001	0.069
	NoSpherA2 - O/Tonto	1.000	0.00001	0.00002	0.100
	NoSpherA2 – O/T	0.999	0.00001	0.00001	0.062
150 K	Tonto/Neutron	0.891	0.00118	0.00169	2.252
	NoSpherA2 - T/Neutron	0.891	0.00118	0.00168	2.245
	NoSpherA2 - O/Neutron	0.890	0.00118	0.00169	2.246
	IAM/Neutron	0.938	0.00121	0.00149	2.054
	NoSpherA2 - T/Tonto	1.000	0.00001	0.00001	0.048
	NoSpherA2 - O/Tonto	1.000	0.00001	0.00001	0.078
	NoSpherA2 – O/T	1.000	0.00001	0.00001	0.064
295 K	Tonto/Neutron	0.965	0.00133	0.00150	1.472
	NoSpherA2 - T/Neutron	0.966	0.00133	0.00150	1.466
	NoSpherA2 - O/Neutron	0.965	0.00133	0.00150	1.465
	IAM/Neutron	1.013	0.00148	0.00178	1.456
	NoSpherA2 - T/Tonto	1.001	0.00001	0.00002	0.061
	NoSpherA2 - O/Tonto	1.001	0.00002	0.00002	0.074
	NoSpherA2 – O/T	1.000	0.00001	0.00001	0.052

Table S14: Statistical results for hydrogen atom ADPs for *L*-Ala, units of differences are Å^2.

Temp.	Method (A/B)	$< U^{ii}_{A} / U^{ii}_{B} >$	$< \mid U^{ij}_{A} - U^{ij}_{B} \mid >$	$< \mid U^{ii}_{A} - U^{ii}_{B} \mid >$	wRMSD
23 K	Tonto/Neutron	1.326	0.0063	0.0070	1.959
	NoSpherA2 - T/Neutron	1.323	0.0062	0.0069	1.938
	NoSpherA2 - O/Neutron	1.319	0.0062	0.0069	1.930
	NoSpherA2 - T/HAR	0.998	0.0002	0.0002	0.048
	NoSpherA2 - O/HAR	0.996	0.0003	0.0003	0.062
	NoSpherA2 – O/T	0.997	0.0001	0.0002	0.040
100 K	HAR/Neutron	--	--	--	--
	NoSpherA2 - T/Neutron	--	--	--	--
	NoSpherA2 - O/Neutron	--	--	--	--
	NoSpherA2 - T/HAR	1.000	0.0003	0.0004	0.125
	NoSpherA2 - O/HAR	0.998	0.0004	0.0004	0.140
	NoSpherA2 – O/T	0.998	0.0001	0.0001	0.054
150 K	HAR/Neutron	1.043	0.0048	0.0056	1.289
	NoSpherA2 - T/Neutron	1.042	0.0046	0.0054	1.273
	NoSpherA2 - O/Neutron	1.040	0.0046	0.0054	1.272
	NoSpherA2 - T/HAR	1.001	0.0006	0.0009	0.123
	NoSpherA2 - O/HAR	0.999	0.0006	0.0009	0.127
	NoSpherA2 – O/T	0.998	0.0001	0.0001	0.026

Table S15: Statistical results for non-hydrogen atom ADPs for *L*-Ala, units of differences are Å^2.

Temp.	Method (A/B)	$< U^{ii}_A / U^{ii}_B >$	$< \| U^{ij}_A - U^{ij}_B \| >$	$< \| U^{ii}_A - U^{ii}_B \| >$	wRMSD
23 K	HAR/Neutron	0.912	0.00034	0.00058	2.830
	NoSpherA2 - T/Neutron	0.915	0.00034	0.00057	2.783
	NoSpherA2 - O/Neutron	0.915	0.00034	0.00057	2.782
	IAM/Neutron	0.973	0.00032	0.00041	1.850
	NoSpherA2 - T/HAR	1.003	0.00001	0.00001	0.117
	NoSpherA2 - O/HAR	1.003	0.00001	0.00001	0.116
	NoSpherA2 – O/T	1.000	0.00000	0.00000	0.016
100 K	HAR/Neutron	--	--	--	--
	NoSpherA2 - T/Neutron	--	--	--	--
	NoSpherA2 - O/Neutron	--	--	--	--
	IAM/Neutron	--	--	--	--
	NoSpherA2 - T/HAR	1.001	0.00001	0.00001	0.135
	NoSpherA2 - O/HAR	1.001	0.00001	0.00001	0.132
	NoSpherA2 – O/T	1.000	0.00000	0.00001	0.086
150 K	HAR/Neutron	0.805	0.00162	0.00296	8.415
	NoSpherA2 - T/Neutron	0.806	0.00161	0.00293	8.374
	NoSpherA2 - O/Neutron	0.806	0.00161	0.00294	8.382
	IAM/Neutron	0.823	0.00154	0.00269	7.132
	NoSpherA2 - T/HAR	1.002	0.00002	0.00002	0.139
	NoSpherA2 - O/HAR	1.001	0.00001	0.00002	0.129
	NoSpherA2 – O/T	1.000	0.00000	0.00001	0.042

Since NoSpherA2-HAR using ORCA as wavefunction source and Tonto-HAR agree very well, for further comparisons ORCA will be used. This second set of comparisons incorporates updated weighting schemes as one variable, and the use of independent software to calculate .tsc files on the other hand.

Table S16: Bond-length statistics for X-H distances, averaged among all bonds for Gly-*L*-Ala.

Temp.	**Method (A / B)**	$\langle r_A/r_B \rangle$	$\langle \lvert \Delta r_{\mathrm{A\text{-}B}} \rvert \rangle$ / Å	**wRMSD**
12K	Tonto / Neutron	0.995	0.0177	1.304
	NoSpherA2 – w / Neutron	0.996	0.0181	1.152
	NoSpherA2 – Q / Neutron	0.995	0.0178	1.296
	NoSpherA2 - w / Tonto	1.000	0.001	0.110
	NoSpherA2 - Q / Tonto	1.000	0.0003	0.055
50K	Tonto / Neutron	1.003	0.010	1.240
	NoSpherA2 – w / Neutron	1.003	0.010	1.191
	NoSpherA2 – Q / Neutron	1.003	0.010	1.176
	NoSpherA2 - w / Tonto	1.000	0.002	0.179
	NoSpherA2 - Q / Tonto	1.000	0.001	0.079
100K	Tonto / Neutron	-	-	-
	NoSpherA2 – w / Neutron	-	-	-
	NoSpherA2 – Q / Neutron	-	-	-
	NoSpherA2 - w / Tonto	1.002	0.002	0.256
	NoSpherA2 - Q / Tonto	1.000	0.001	0.135
150K	Tonto / Neutron	1.000	0.009	1.375
	NoSpherA2 – w / Neutron	1.000	0.010	1.419
	NoSpherA2 – Q / Neutron	1.000	0.009	1.376
	NoSpherA2 - w / Tonto	1.001	0.001	0.179
	NoSpherA2 - Q / Tonto	1.000	0.000	0.064
295K	Tonto / Neutron	0.994	0.011	1.269
	NoSpherA2 – w / Neutron	0.995	0.012	1.300
	NoSpherA2 – Q / Neutron	0.994	0.011	1.275
	NoSpherA2 - w / Tonto	1.001	0.003	0.237
	NoSpherA2 - Q / Tonto	1.000	0.001	0.051

Table S17: Bond-length statistics for X-H distances, averaged among all bonds for *L*-Ala.

Temp.	Method (A / B)	$\langle r_A/r_B \rangle$	$\langle \lvert \Delta r_{\text{A-B}} \rvert \rangle$ / Å	wRMSD
23K	Tonto / Neutron	0.991	0.011	1.904
	NoSpherA2 – w / Neutron	0.991	0.011	1.838
	NoSpherA2 – Q / Neutron	0.991	0.010	1.896
	Tonto / NoSpherA2 – w	1.001	0.0009	0.123
	Tonto / NoSpherA2 – Q	1.000	0.0006	0.086
100K	Tonto / Neutron	-	-	-
	NoSpherA2 – w / Neutron	-	-	-
	NoSpherA2 – Q / Neutron	-	-	-
	Tonto / NoSpherA2 – w	0.999	0.0033	0.650
	Tonto / NoSpherA2 – Q	1.000	0.0003	0.094
150K	Tonto / Neutron	0.995	0.008	1.197
	NoSpherA2 – w / Neutron	0.996	0.012	1.395
	NoSpherA2 – Q / Neutron	0.995	0.008	1.081
	Tonto / NoSpherA2 – w	1.001	0.0049	0.458
	Tonto / NoSpherA2 – Q	1.001	0.0010	0.102

Table S18: Statistical results for hydrogen atom ADPs for Gly-*L*-Ala, units of differences are Å^2.

Temp.	Method (A/B)	$< U^{ii}_A / U^{ii}_B >$	$< \| U^{ij}_A - U^{ij}_B \| >$	$< \| U^{ii}_A - U^{ii}_B \| >$	wRMSD
12 K	Tonto/Neutron	1.058	0.005	0.005	1.382
	NoSpherA2 - w/Neutron	1.051	0.0048	0.0052	1.305
	NoSpherA2 - Q/Neutron	1.054	0.0050	0.0054	1.395
	NoSpherA2 - w/Tonto	0.997	0.0008	0.0009	0.188
	NoSpherA2 - Q/Tonto	0.996	0.0003	0.0003	0.071
50 K	Tonto/Neutron	1.024	0.0054	0.0067	1.359
	NoSpherA2 - w/Neutron	1.011	0.0054	0.0068	1.298
	NoSpherA2 - Q/Neutron	1.018	0.0056	0.0071	1.385
	NoSpherA2 - w/Tonto	0.985	0.0012	0.0011	0.230
	NoSpherA2 - Q/Tonto	0.991	0.0005	0.0006	0.107
100 K	Tonto/Neutron	-	-	-	-
	NoSpherA2 - w/Neutron	-	-	-	-
	NoSpherA2 - Q/Neutron	-	-	-	-
	NoSpherA2 - w/Tonto	1.006	0.0018	0.0022	0.375
	NoSpherA2 - Q/Tonto	0.996	0.0005	0.0006	0.096
150 K	Tonto/Neutron	1.059	0.0051	0.0052	1.316
	NoSpherA2 - w/Neutron	1.055	0.0049	0.0051	1.216
	NoSpherA2 - Q/Neutron	1.056	0.0051	0.0053	1.318
	NoSpherA2 - w/Tonto	0.999	0.0009	0.0010	0.195
	NoSpherA2 - Q/Tonto	0.997	0.0003	0.0003	0.061
295 K	Tonto/Neutron	1.117	0.0098	0.0107	1.623
	NoSpherA2 - w/Neutron	1.109	0.0099	0.0106	1.588
	NoSpherA2 - Q/Neutron	1.117	0.0100	0.0109	1.637
	NoSpherA2 - w/Tonto	0.995	0.0016	0.0019	0.364
	NoSpherA2 - Q/Tonto	0.999	0.0006	0.0006	0.083

Table S19: Statistical results for non-hydrogen atom ADPs for Gly-*L*-Ala, units of differences are Å^2.

Temp.	**Method (A/B)**	$< U^{ii}_{A} / U^{ii}_{B} >$	$< \| U^{ij}_{A} - U^{ij}_{B} \| >$	$< \| U^{ii}_{A} - U^{ii}_{B} \| >$	**wRMSD**
12 K	Tonto/Neutron	0.748	0.001	0.002	2.586
	NoSpherA2 - w/Neutron	0.760	0.0011	0.0016	2.529
	NoSpherA2 - Q/Neutron	0.748	0.0011	0.0017	2.581
	NoSpherA2 - w/Tonto	1.016	0.0000	0.0001	0.315
	NoSpherA2 - Q/Tonto	1.000	0.0000	0.0000	0.080
50 K	Tonto/Neutron	0.783	0.0012	0.0018	2.675
	NoSpherA2 - w/Neutron	0.792	0.0012	0.0017	2.611
	NoSpherA2 - Q/Neutron	0.783	0.0012	0.0018	2.674
	NoSpherA2 - w/Tonto	1.013	0.0000	0.0001	0.277
	NoSpherA2 - Q/Tonto	1.000	0.0000	0.0000	0.088
100 K	Tonto/Neutron	-	-	-	-
	NoSpherA2 - w/Neutron	-	-	-	-
	NoSpherA2 - Q/Neutron	-	-	-	-
	NoSpherA2 - w/Tonto	1.013	0.0001	0.0001	0.688
	NoSpherA2 - Q/Tonto	1.000	0.0000	0.0000	0.103
150 K	Tonto/Neutron	0.891	0.0012	0.0017	2.252
	NoSpherA2 - w/Neutron	0.897	0.0011	0.0016	2.194
	NoSpherA2 - Q/Neutron	0.890	0.0012	0.0017	2.247
	NoSpherA2 - w/Tonto	1.007	0.0001	0.0001	0.369
	NoSpherA2 - Q/Tonto	1.000	0.0000	0.0000	0.085
295 K	Tonto/Neutron	0.965	0.0013	0.0015	1.472
	NoSpherA2 - w/Neutron	0.969	0.0013	0.0015	1.446
	NoSpherA2 - Q/Neutron	0.965	0.0013	0.0015	1.465
	NoSpherA2 - w/Tonto	1.005	0.0001	0.0001	0.364
	NoSpherA2 - Q/Tonto	1.001	0.0000	0.0000	0.074

Table S20: Statistical results for hydrogen atom ADPs for *L*-Ala, units of differences are Å^2.

Temp.	**Method (A/B)**	**< U^{ii}_A / U^{ii}_B >**	**< | $U^{ij}_A - U^{ij}_B$ | >**	**< | $U^{ii}_A - U^{ii}_B$ | >**	**wRMSD**
23 K	Tonto/Neutron	1.326	0.0063	0.0070	1.959
	NoSpherA2 - w/Neutron	1.302	0.0063	0.0066	1.876
	NoSpherA2 - Q/Neutron	1.336	0.0063	0.0071	1.893
	NoSpherA2 - w/Tonto	0.983	0.0006	0.0006	0.118
	NoSpherA2 - Q/Tonto	1.006	0.0004	0.0004	0.079
100 K	Tonto/Neutron	-	-	-	-
	NoSpherA2 - w/Neutron	-	-	-	-
	NoSpherA2 - Q/Neutron	-	-	-	-
	NoSpherA2 - w/Tonto	0.981	0.0017	0.0020	0.751
	NoSpherA2 - Q/Tonto	1.002	0.0004	0.0004	0.154
150 K	Tonto/Neutron	1.043	0.0048	0.0056	1.289
	NoSpherA2 - w/Neutron	1.022	0.0046	0.0047	1.177
	NoSpherA2 - Q/Neutron	1.045	0.0046	0.0054	1.272
	NoSpherA2 - w/Tonto	0.993	0.0017	0.0023	0.346
	NoSpherA2 - Q/Tonto	1.003	0.0007	0.0010	0.140

Table S21: Statistical results for non-hydrogen atom ADPs for *L*-Ala, units of differences are Å^2.

Temp.	**Method (A/B)**	$\langle U^{ii}_A / U^{ii}_B \rangle$	$\langle \lvert U^{ij}_A - U^{ij}_B \rvert \rangle$	$\langle \lvert U^{ii}_A - U^{ii}_B \rvert \rangle$	**wRMSD**
23 K	Tonto/Neutron	0.912	0.000	0.001	2.830
	NoSpherA2 - w/Neutron	0.916	0.00033	0.00057	2.771
	NoSpherA2 - Q/Neutron	0.915	0.00000	0.00057	2.764
	NoSpherA2 - w/Tonto	1.004	0.00001	0.00002	0.148
	NoSpherA2 - Q/Tonto	1.003	0.00002	0.00001	0.167
100 K	Tonto/Neutron	-	-	-	-
	NoSpherA2 - w/Neutron	-	-	-	-
	NoSpherA2 - Q/Neutron	-	-	-	-
	NoSpherA2 - w/Tonto	1.002	0.00004	0.00004	0.783
	NoSpherA2 - Q/Tonto	1.000	0.00001	0.00001	0.156
150 K	Tonto/Neutron	0.805	0.00162	0.00296	8.415
	NoSpherA2 - w/Neutron	0.807	0.00160	0.00293	8.376
	NoSpherA2 - Q/Neutron	0.806	0.00013	0.00294	8.372
	NoSpherA2 - w/Tonto	1.003	0.00004	0.00005	0.368
	NoSpherA2 - Q/Tonto	1.001	0.00002	0.00002	0.153

The refinements and data analysis were repeated *without* the I/σ(I) cutoff to see the difference this makes although inside *olex2.refine* it is normally applied automatically. Most datasets will not have been pruned in such a manner; therefore, the following tables correspond to the situation for the usual data. For comparison of the implementation and numerical reproducibility, the first set of data tables above is more important, but to understand the effect of pruning and cutoff criteria, the second set of table provides additional insight.

Table S22: Bond-length statistics for X-H distances, averaged among all bonds for Gly-*L*-Ala.

Temp.	Method (A / B)	$\langle r_A/r_B \rangle$	$\langle \lvert \Delta r_{\text{A-B}} \rvert \rangle$ / Å	wRMSD
12K	Tonto / Neutron	0.996	0.006	1.295
	NoSpherA2 – T / Neutron	0.996	0.007	1.304
	NoSpherA2 – O / Neutron	0.996	0.006	1.268
	IAM / Neutron	0.871	0.139	11.588
	Tonto / NoSpherA2 – T	1.000	0.0003	0.056
	Tonto / NoSpherA2 – O	1.000	0.0004	0.074
	NoSpherA2 – O / NoSpherA2 – T	1.000	0.0003	0.060
50K	Tonto / Neutron	1.003	0.010	1.236
	NoSpherA2 – T / Neutron	1.004	0.010	1.187
	NoSpherA2 – O / Neutron	1.003	0.010	1.177
	IAM / Neutron	0.883	0.125	12.807
	Tonto / NoSpherA2 – T	1.000	0.001	0.077
	Tonto / NoSpherA2 – O	0.999	0.001	0.101
	NoSpherA2 – O / NoSpherA2 – T	1.000	0.001	0.056
100K	Tonto / Neutron	-	-	-
	NoSpherA2 – T / Neutron	-	-	-
	NoSpherA2 – O / Neutron	-	-	-
	IAM / Neutron	-	-	-
	Tonto / NoSpherA2 – T	1.000	0.001	0.097
	Tonto / NoSpherA2 – O	1.000	0.002	0.290
	NoSpherA2 – O / NoSpherA2 – T	1.000	0.002	0.300
150K	Tonto / Neutron	1.000	0.008	1.378
	NoSpherA2 – T / Neutron	1.000	0.009	1.416
	NoSpherA2 – O / Neutron	1.000	0.009	1.384
	IAM / Neutron	0.880	0.128	10.767
	Tonto / NoSpherA2 – T	1.000	0.001	0.106
	Tonto / NoSpherA2 – O	1.000	0.001	0.094
	NoSpherA2 – O / NoSpherA2 – T	1.000	0.0003	0.055
295K	Tonto / Neutron	0.994	0.011	1.061
	NoSpherA2 – T / Neutron	0.994	0.011	1.092
	NoSpherA2 – O / Neutron	0.994	0.011	1.105
	IAM / Neutron	0.887	0.120	9.113
	Tonto / NoSpherA2 – T	1.000	0.001	0.101
	Tonto / NoSpherA2 – O	1.000	0.001	0.094
	NoSpherA2 – O / NoSpherA2 – T	1.000	0.0003	0.051

Table S23: Bond-length statistics for X-H distances, averaged among all bonds for *L*-Ala.

Temp.	**Method (A / B)**	$\langle r_A/r_B \rangle$	$\langle\lvert\Delta r_{\text{A-B}}\rvert\rangle$ / Å	**wRMSD**
23K	Tonto / Neutron	0.991	0.010	1.875
	NoSpherA2 – T / Neutron	0.991	0.010	1.789
	NoSpherA2 – O / Neutron	0.991	0.010	1.841
	IAM / Neutron	0.916	0.090	9.070
	Tonto / NoSpherA2 – T	1.000	0.0009	0.109
	Tonto / NoSpherA2 – O	1.000	0.0007	0.086
	NoSpherA2 – O / NoSpherA2 – T	1.000	0.0004	0.066
100K	Tonto / Neutron	-	-	-
	NoSpherA2 – T / Neutron	-	-	-
	NoSpherA2 – O / Neutron	-	-	-
	IAM / Neutron	-	-	-
	Tonto / NoSpherA2 – T	1.000	0.0003	0.094
	Tonto / NoSpherA2 – O	1.000	0.0004	0.116
	NoSpherA2 – O / NoSpherA2 – T	1.000	0.0004	0.116
150K	Tonto / Neutron	0.995	0.009	1.186
	NoSpherA2 – T / Neutron	0.996	0.007	0.943
	NoSpherA2 – O / Neutron	0.992	0.007	0.974
	IAM / Neutron	0.902	0.105	9.733
	Tonto / NoSpherA2 – T	0.999	0.0036	0.342
	Tonto / NoSpherA2 – O	0.999	0.0034	0.337
	NoSpherA2 – O / NoSpherA2 – T	1.000	0.0001	0.024

Table S24: Statistical results for hydrogen atom ADPs for Gly-*L*-Ala, units of differences are Å^2.

Temp.	Method (A/B)	wRMSD	$< \| U^{ij}_A - U^{ij}_B \| >$	$< \| U^{ii}_A - U^{ii}_B \| >$	$< U^{ii}_A / U^{ii}_B >$
12 K	Tonto/Neutron	1.363	0.0048	0.0053	1.060
	NoSpherA2 - T/Neutron	1.386	0.0049	0.0053	1.056
	NoSpherA2 - O/Neutron	1.387	0.0049	0.0053	1.056
	NoSpherA2 - T/Tonto	0.071	0.0003	0.0004	0.996
	NoSpherA2 - O/Tonto	0.083	0.0003	0.0004	0.996
	NoSpherA2 - O/T	0.026	0.0001	0.0001	1.000
50 K	Tonto/Neutron	1.342	0.0053	0.0067	1.021
	NoSpherA2 - T/Neutron	1.387	0.0055	0.0069	1.015
	NoSpherA2 - O/Neutron	1.375	0.0055	0.0069	1.014
	NoSpherA2 - T/Tonto	0.092	0.0005	0.0006	0.994
	NoSpherA2 - O/Tonto	0.096	0.0005	0.0006	0.993
	NoSpherA2 - O/T	0.023	0.0001	0.0001	0.999
100 K	Tonto/Neutron	-	-	-	-
	NoSpherA2 - T/Neutron	-	-	-	-
	NoSpherA2 - O/Neutron	-	-	-	-
	NoSpherA2 - T/Tonto	0.079	0.0003	0.0004	0.994
	NoSpherA2 - O/Tonto	0.081	0.0003	0.0004	0.994
	NoSpherA2 - O/T	0.027	0.0001	0.0001	1.000
150 K	Tonto/Neutron	1.313	0.0078	0.0051	1.056
	NoSpherA2 - T/Neutron	1.327	0.0050	0.0051	1.053
	NoSpherA2 - O/Neutron	1.321	0.0050	0.0051	1.053
	NoSpherA2 - T/Tonto	0.070	0.0003	0.0003	0.997
	NoSpherA2 - O/Tonto	0.069	0.0003	0.0004	0.997
	NoSpherA2 - O/T	0.026	0.0001	0.0001	1.000
295 K	Tonto/Neutron	1.629	0.0096	0.0106	1.116
	NoSpherA2 - T/Neutron	1.679	0.0099	0.0108	1.113
	NoSpherA2 - O/Neutron	1.678	0.0099	0.0108	1.113
	NoSpherA2 - T/Tonto	0.088	0.0006	0.0006	0.997
	NoSpherA2 - O/Tonto	0.090	0.0006	0.0006	0.996
	NoSpherA2 - O/T	0.027	0.0001	0.0001	1.000

Table S25: Statistical results for non-hydrogen atom ADPs for Gly-*L*-Ala, units of differences are Å^2.

Temp.	Method (A/B)	wRMSD	$< \vert U^{ij}_A - U^{ij}_B \vert >$	$< \vert U^{ii}_A - U^{ii}_B \vert >$	$< U^{ii}_A / U^{ii}_B >$
12 K	IAM/Neutron	2.273	0.00108	0.00140	0.863
	Tonto/Neutron	2.578	0.00110	0.00167	0.749
	NoSpherA2 - T/Neutron	2.577	0.00110	0.00168	0.749
	NoSpherA2 - O/Neutron	2.579	0.00111	0.00168	0.749
	NoSpherA2 - T/Tonto	0.057	0.00001	0.00001	1.000
	NoSpherA2 - O/Tonto	0.078	0.00001	0.00001	0.999
	NoSpherA2 - O/T	0.061	0.00001	0.00001	0.999
50 K	IAM/Neutron	2.212	0.00107	0.00144	0.889
	Tonto/Neutron	2.669	0.00122	0.00178	0.784
	NoSpherA2 - T/Neutron	2.666	0.00121	0.00177	0.785
	NoSpherA2 - O/Neutron	2.667	0.00122	0.00178	0.785
	NoSpherA2 - T/Tonto	0.085	0.00001	0.00002	1.001
	NoSpherA2 - O/Tonto	0.101	0.00002	0.00002	1.000
	NoSpherA2 - O/T	0.051	0.00001	0.00001	0.999
100 K	IAM/Neutron	-	-	-	-
	Tonto/Neutron	-	-	-	-
	NoSpherA2 - T/Neutron	-	-	-	-
	NoSpherA2 - O/Neutron	-	-	-	-
	NoSpherA2 - T/Tonto	0.080	0.00001	0.00002	1.000
	NoSpherA2 - O/Tonto	0.103	0.00001	0.00002	1.000
	NoSpherA2 - O/T	0.064	0.00001	0.00001	0.999
150 K	IAM/Neutron	2.068	0.00122	0.00149	0.937
	Tonto/Neutron	2.251	0.00118	0.00168	0.891
	NoSpherA2 - T/Neutron	2.268	0.00118	0.00168	0.891
	NoSpherA2 - O/Neutron	2.271	0.00118	0.00169	0.891
	NoSpherA2 - T/Tonto	0.061	0.00001	0.00001	1.000
	NoSpherA2 - O/Tonto	0.087	0.00001	0.00002	1.000
	NoSpherA2 - O/T	0.065	0.00001	0.00001	1.000
295 K	IAM/Neutron	1.468	0.00147	0.00177	1.011
	Tonto/Neutron	1.468	0.00133	0.00149	0.966
	NoSpherA2 - T/Neutron	1.465	0.00133	0.00149	0.966
	NoSpherA2 - O/Neutron	1.464	0.00133	0.00149	0.966
	NoSpherA2 - T/Tonto	0.069	0.00002	0.00002	1.001
	NoSpherA2 - O/Tonto	0.081	0.00002	0.00002	1.000
	NoSpherA2 - O/T	0.054	0.00001	0.00001	1.000

Table S26: Statistical results for hydrogen atom ADPs for *L*-Ala, units of differences are Å^2.

Temp.	**Method (A/B)**	**wRMSD**	**$< \mid U^{ij}_A - U^{ij}_B \mid >$**	**$< \mid U^{ii}_A - U^{ii}_B \mid >$**	**$< U^{ii}_A / U^{ii}_B >$**
23 K	Tonto/Neutron	1.957	0.0063	0.0069	1.328
	NoSpherA2 - T/Neutron	1.979	0.0062	0.0069	1.328
	NoSpherA2 - O/Neutron	1.971	0.0061	0.0068	1.324
	NoSpherA2 - T/Tonto	0.141	0.0005	0.0006	1.001
	NoSpherA2 - O/Tonto	0.154	0.0006	0.0007	0.998
	NoSpherA2 - O/T	0.029	0.0002	0.0001	0.997
100 K	Tonto/Neutron	-	-	-	-
	NoSpherA2 - T/Neutron	-	-	-	-
	NoSpherA2 - O/Neutron	-	-	-	-
	NoSpherA2 - T/Tonto	0.182	0.0003	0.0005	1.001
	NoSpherA2 - O/Tonto	0.190	0.0005	0.0005	1.000
	NoSpherA2 - O/T	0.051	0.0001	0.0001	0.998
150 K	Tonto/Neutron	1.282	0.0055	0.0056	1.042
	NoSpherA2 - T/Neutron	1.252	0.0050	0.0058	1.074
	NoSpherA2 - O/Neutron	1.251	0.0050	0.0058	1.072
	NoSpherA2 - T/Tonto	0.513	0.0028	0.0026	1.031
	NoSpherA2 - O/Tonto	0.511	0.0027	0.0026	1.028
	NoSpherA2 - O/T	0.023	0.0001	0.0002	0.998

Table S27: Statistical results for non-hydrogen atom ADPs for *L*-Ala, units of differences are Å^2.

Temp.	Method (A/B)	wRMSD	$< \mid U^{ij}_A - U^{ij}_B \mid >$	$< \mid U^{ii}_A - U^{ii}_B \mid >$	$< U^{ii}_A / U^{ii}_B >$
23 K	IAM/Neutron	3.667	0.00032	0.00043	0.967
	Tonto/Neutron	2.818	0.00034	0.00058	0.913
	NoSpherA2 - T/Neutron	2.802	0.00034	0.00057	0.915
	NoSpherA2 - O/Neutron	2.800	0.00034	0.00057	0.915
	NoSpherA2 - T/Tonto	0.136	0.00001	0.00001	1.002
	NoSpherA2 - O/Tonto	0.133	0.00001	0.00001	1.002
	NoSpherA2 - O/T	0.029	0.000002	0.000003	1.000
100 K	IAM/Neutron	-	-	-	-
	Tonto/Neutron	-	-	-	-
	NoSpherA2 - T/Neutron	-	-	-	-
	NoSpherA2 - O/Neutron	-	-	-	-
	NoSpherA2 - T/Tonto	0.432	0.00003	0.00003	1.001
	NoSpherA2 - O/Tonto	0.425	0.00003	0.00003	1.001
	NoSpherA2 - O/T	0.081	0.000005	0.000007	1.000
150 K	IAM/Neutron	51.759	0.00156	0.00272	0.821
	Tonto/Neutron	8.415	0.00162	0.00295	0.805
	NoSpherA2 - T/Neutron	8.120	0.00162	0.00292	0.807
	NoSpherA2 - O/Neutron	8.127	0.00162	0.00292	0.807
	NoSpherA2 - T/Tonto	1.420	0.0002	0.00013	1.003
	NoSpherA2 - O/Tonto	1.421	0.0002	0.00014	1.003
	NoSpherA2 - O/T	0.033	0.000004	0.000006	1.000

Table S28: Bond-length statistics for X-H distances, averaged among all bonds for Gly-*L*-Ala.

Temp.	**Method (A / B)**	$\langle r_A/r_B \rangle$	$\langle\lvert\Delta r_{\text{A-B}}\rvert\rangle$ / Å	**wRMSD**
12K	Tonto / Neutron	0.996	0.006	1.295
	NoSpherA2 – w / Neutron	0.996	0.006	1.137
	NoSpherA2 – Q / Neutron	0.996	0.007	1.286
	Tonto / NoSpherA2 – w	1.000	0.0008	0.101
	Tonto / NoSpherA2 – Q	1.000	0.0008	0.117
50K	Tonto / Neutron	1.003	0.010	1.236
	NoSpherA2 – w / Neutron	1.003	0.010	1.189
	NoSpherA2 – Q / Neutron	1.003	0.010	1.209
	Tonto / NoSpherA2 – w	1.000	0.0014	0.141
	Tonto / NoSpherA2 – Q	1.000	0.0003	0.043
100K	Tonto / Neutron	-	-	-
	NoSpherA2 – w / Neutron	-	-	-
	NoSpherA2 – Q / Neutron	-	-	-
	Tonto / NoSpherA2 – w	0.999	0.0022	0.249
	Tonto / NoSpherA2 – Q	1.000	0.0009	0.118
150K	Tonto / Neutron	1.000	0.008	1.378
	NoSpherA2 – w / Neutron	1.000	0.010	1.478
	NoSpherA2 – Q / Neutron	1.000	0.009	1.390
	Tonto / NoSpherA2 – w	1.000	0.0013	0.176
	Tonto / NoSpherA2 – Q	1.000	0.0008	0.111
295K	Tonto / Neutron	0.994	0.011	1.061
	NoSpherA2 – w / Neutron	0.994	0.011	1.092
	NoSpherA2 – Q / Neutron	0.994	0.011	1.048
	Tonto / NoSpherA2 – w	1.000	0.0011	0.101
	Tonto / NoSpherA2 – Q	1.000	0.0008	0.083

Table S29: Bond-length statistics for X-H distances, averaged among all bonds for *L*-Ala.

Temp.	Method (A / B)	$\langle r_A/r_B \rangle$	$\langle \lvert \Delta r_{\text{A-B}} \rvert \rangle$ / Å	wRMSD
23K	Tonto / Neutron	0.991	0.010	1.875
	NoSpherA2 – w / Neutron	0.991	0.011	1.835
	NoSpherA2 – Q / Neutron	0.991	0.010	1.729
	Tonto / NoSpherA2 – w	1.000	0.0013	0.172
	Tonto / NoSpherA2 – Q	1.000	0.0009	0.089
100K	Tonto / Neutron	-	-	-
	NoSpherA2 – w / Neutron	-	-	-
	NoSpherA2 – Q / Neutron	-	-	-
	Tonto / NoSpherA2 – w	1.002	0.0033	0.680
	Tonto / NoSpherA2 – Q	1.001	0.0007	0.186
150K	Tonto / Neutron	0.995	0.009	1.186
	NoSpherA2 – w / Neutron	0.996	0.007	0.991
	NoSpherA2 – Q / Neutron	0.996	0.007	0.966
	Tonto / NoSpherA2 – w	0.998	0.0033	0.295
	Tonto / NoSpherA2 – Q	0.999	0.0039	0.358

Table S30: Statistical results for hydrogen atom ADPs for Gly-*L*-Ala, units of differences are Å^2.

Temp.	Method (A/B)	wRMSD	$< \mid U^{ij}_A - U^{ij}_B \mid >$	$< \mid U^{ii}_A - U^{ii}_B \mid >$	$< U^{ii}_A / U^{ii}_B >$
12 K	Tonto/Neutron	1.363	0.0048	0.0053	1.060
	NoSpherA2 - w/Neutron	1.308	0.0048	0.0052	1.051
	NoSpherA2 - Q/Neutron	1.382	0.0049	0.0052	1.062
	NoSpherA2 - w/Tonto	0.221	0.0010	0.0011	0.995
	NoSpherA2 - Q/Tonto	0.113	0.0005	0.0006	1.003
50 K	Tonto/Neutron	1.342	0.0053	0.0067	1.021
	NoSpherA2 - w/Neutron	1.303	0.0054	0.0068	1.011
	NoSpherA2 - Q/Neutron	1.369	0.0056	0.0069	1.021
	NoSpherA2 - w/Tonto	0.226	0.0011	0.0012	0.989
	NoSpherA2 - Q/Tonto	0.113	0.0006	0.0007	1.002
100 K	Tonto/Neutron	-	-	-	-
	NoSpherA2 - w/Neutron	-	-	-	-
	NoSpherA2 - Q/Neutron	-	-	-	-
	NoSpherA2 - w/Tonto	0.384	0.0019	0.0021	1.000
	NoSpherA2 - Q/Tonto	0.127	0.0005	0.0006	1.002
150 K	Tonto/Neutron	1.313	0.0078	0.0051	1.056
	NoSpherA2 - w/Neutron	1.235	0.0049	0.0050	1.049
	NoSpherA2 - Q/Neutron	1.323	0.0050	0.0051	1.060
	NoSpherA2 - w/Tonto	0.203	0.0009	0.0010	0.995
	NoSpherA2 - Q/Tonto	0.100	0.0005	0.0005	1.003
295 K	Tonto/Neutron	1.629	0.0096	0.0106	1.116
	NoSpherA2 - w/Neutron	1.623	0.0097	0.0105	1.106
	NoSpherA2 - Q/Neutron	1.679	0.0010	0.0011	1.119
	NoSpherA2 - w/Tonto	0.238	0.0016	0.0018	0.992
	NoSpherA2 - Q/Tonto	0.102	0.0006	0.0006	1.002

Table S31: Statistical results for non-hydrogen atom ADPs for Gly-*L*-Ala, units of differences are Å^2.

Temp.	**Method (A/B)**	**wRMSD**	$\langle \vert U^{ij}_A - U^{ij}_B \vert \rangle$	$\langle \vert U^{ii}_A - U^{ii}_B \vert \rangle$	$\langle U^{ii}_A / U^{ii}_B \rangle$
12 K	Tonto/Neutron	2.578	0.00110	0.00167	0.749
	NoSpherA2 - w/Neutron	2.520	0.00108	0.00163	0.762
	NoSpherA2 - Q/Neutron	2.575	0.00110	0.00167	0.749
	NoSpherA2 - w/Tonto	0.336	0.00005	0.00007	1.017
	NoSpherA2 - Q/Tonto	0.144	0.00002	0.00002	1.000
50 K	Tonto/Neutron	2.669	0.00122	0.00178	0.784
	NoSpherA2 - w/Neutron	2.594	0.00118	0.00171	0.795
	NoSpherA2 - Q/Neutron	2.665	0.00121	0.00178	0.785
	NoSpherA2 - w/Tonto	0.325	0.00006	0.00008	1.015
	NoSpherA2 - Q/Tonto	0.152	0.00002	0.00003	1.000
100 K	Tonto/Neutron	-	-	-	-
	NoSpherA2 - w/Neutron	-	-	-	-
	NoSpherA2 - Q/Neutron	-	-	-	-
	NoSpherA2 - w/Tonto	0.745	0.00009	0.00013	1.014
	NoSpherA2 - Q/Tonto	0.156	0.00002	0.00002	1.000
150 K	Tonto/Neutron	2.251	0.00118	0.00168	0.891
	NoSpherA2 - w/Neutron	2.226	0.00114	0.00161	0.897
	NoSpherA2 - Q/Neutron	2.270	0.00117	0.00168	0.891
	NoSpherA2 - w/Tonto	0.400	0.00006	0.00009	1.007
	NoSpherA2 - Q/Tonto	0.130	0.00002	0.00002	1.000
295 K	Tonto/Neutron	1.468	0.00133	0.00149	0.966
	NoSpherA2 - w/Neutron	1.445	0.00132	0.00147	0.970
	NoSpherA2 - Q/Neutron	1.464	0.00133	0.00150	0.966
	NoSpherA2 - w/Tonto	0.409	0.00008	0.00012	1.005
	NoSpherA2 - Q/Tonto	0.103	0.00002	0.00003	1.001

Table S32: Statistical results for hydrogen atom ADPs for *L*-Ala, units of differences are Å^2.

Temp.	**Method (A/B)**	**wRMSD**	**$< \mid U^{ij}_A - U^{ij}_B \mid >$**	**$< \mid U^{ii}_A - U^{ii}_B \mid >$**	**$< U^{ii}_A / U^{ii}_B >$**
23 K	Tonto/Neutron	1.957	0.0063	0.0069	1.328
	NoSpherA2 - w/Neutron	1.925	0.0061	0.0066	1.311
	NoSpherA2 - Q/Neutron	1.826	0.0063	0.0074	1.374
	NoSpherA2 - w/Tonto	0.202	0.0010	0.0010	0.989
	NoSpherA2 - Q/Tonto	0.180	0.0009	0.0011	1.039
100 K	Tonto/Neutron	-	-	-	-
	NoSpherA2 - w/Neutron	-	-	-	-
	NoSpherA2 - Q/Neutron	-	-	-	-
	NoSpherA2 - w/Tonto	0.672	0.0044	0.0020	0.984
	NoSpherA2 - Q/Tonto	0.217	0.0038	0.0007	1.006
150 K	Tonto/Neutron	1.282	0.0055	0.0056	1.042
	NoSpherA2 - w/Neutron	1.230	0.0046	0.0056	1.044
	NoSpherA2 - Q/Neutron	1.179	0.0048	0.0055	1.077
	NoSpherA2 - w/Tonto	0.279	0.0015	0.0017	1.000
	NoSpherA2 - Q/Tonto	0.467	0.0026	0.0026	1.033

Table S33: Statistical results for non-hydrogen atom ADPs for *L*-Ala, units of differences are Å^2.

Temp.	**Method (A/B)**	**wRMSD**	$\langle \lvert U^{ij}_{A} - U^{ij}_{B} \rvert \rangle$	$\langle \lvert U^{ii}_{A} - U^{ii}_{B} \rvert \rangle$	$\langle U^{ii}_{A} / U^{ii}_{B} \rangle$
23 K	Tonto/Neutron	2.818	0.00034	0.00058	0.913
	NoSpherA2 - w/Neutron	2.790	0.00033	0.00057	0.916
	NoSpherA2 - Q/Neutron	2.708	0.00033	0.00057	0.917
	NoSpherA2 - w/Tonto	0.156	0.00001	0.00002	1.003
	NoSpherA2 - Q/Tonto	0.311	0.00003	0.00003	1.004
100 K	Tonto/Neutron	-	-	-	-
	NoSpherA2 - w/Neutron	-	-	-	-
	NoSpherA2 - Q/Neutron	-	-	-	-
	NoSpherA2 - w/Tonto	0.904	0.00005	0.00005	1.003
	NoSpherA2 - Q/Tonto	0.439	0.00003	0.00003	1.001
150 K	Tonto/Neutron	8.415	0.00162	0.00295	0.805
	NoSpherA2 - w/Neutron	8.323	0.00161	0.00295	0.806
	NoSpherA2 - Q/Neutron	8.089	0.00162	0.00292	0.808
	NoSpherA2 - w/Tonto	0.358	0.00004	0.00005	1.001
	NoSpherA2 - Q/Tonto	1.429	0.00017	0.00014	1.003

References

S1. L. J. Bourhis, O. V. Dolomanov, R. J. Gildea, J. A. K. Howard, H. Puschmann, *Acta Cryst.*, 2015, **A71**, 59–75.

S2. R. W Grosse-Kunstleve and P. D. Adams, *J. Appl. Cryst.*, 2002, **35**, 477-480.

S3. O. V. Dolomanov, L. J. Bourhis, R. J. Gildea, J. A. K. Howard, H. Puschmann, J. *Appl. Cryst.*, 2009, **42**, 339–341.

S4. R. F. Stewart, E. R. Davidson, W. T. Simpson, *J. Chem. Phys.*, **1965**, 42(9), 3175–3187.

S5. S. C. Capelli, H.-B. Bürgi, B. Dittrich, S. Grabowsky, D. Jayatilaka, *IUCrJ,* 2014, **1**, 361-379.

S6. R. Destro, R. E. Marsh, R. Bianchi, *J. Phys. Chem.,* 1998, **92**, 966–973.

S7. S. C. Capelli, H.-B. Bürgi, S. A. Mason, D. Jayatilaka, *Acta Cryst. C* **2014**, *70*, 949–952.

S8. L. A. Malaspina, E. K. Wieduwilt, J. Bergmann, F. Kleemiss, B. Meyer, M. F. Ruiz-López, R. Pal, E. Hupf, J. Beckmann, R. O. Piltz, A. J. Edwards, S. Grabowsky, A. Genoni, *J. Phys. Chem. Let.* 2019, **10**, 6973-6982.

B.2 Averaged Interaction Energy Scripts

Listing B.1: *bash* Script to Calculate Interaction Energies on a Series of Geometries.

```
#!/bin/bash

#
#Information about expected job size in PBS script
#
    ncpus_g09=4
    mem_g09=30

#
#manually prepare xyz file with VMD to have trajectory of all wanted residuals individually
#

#
#Give name of ligand file
#Give charge of ligand
#Give number of atoms in ligand
#
    ligand=sila
    charge=-1
    natoms_ligand=32

#
#Give how many fragments we have surrounding central molecule
#Give names of files in which traj is stored for each fragment
#Give corresponding charge (after saturation)
#Give number of atoms (!INCLUDING SATURATION ATOMS!)
#

    number_fragments=13
    fragments=(       arg  glyala  hsd  leuser  leutyr  leu  metval  phe  serleu  trp  tyr  val1  val2)
    charges=(         1    0       0    0       0       0    0       0    0       0    0    0     0)
    natoms_fragments=(24   17      17   30      40      19   33      20   30      24   21   16    16)

#
#Give number of frames in trajectory files
#
    frames=1000

#
#Prepare gaussian input template (change mem and procs if needed)
#
function gaussian_header {
    echo "%mem=20GB" > $1
    echo "%nproc=4" >> $1
    echo "#P␣B3LYP/6-31G(d,p)␣6d␣10f␣NoSymm␣FChk" >> $1
    echo "␣" >> $1
    echo "Job" >> $1
    echo "␣" >> $1
}

#
#function which adds saturating hydrogen to amino acid residuals through z-matrix
#
function saturation {
#$1=number of atoms (N is expected to be first atom in xyz file, CO last atoms)
#$2=com file to write into
    echo "H␣1␣1.07␣3␣120.00␣2␣180.00" >> ${2}
    echo "H␣$((␣${1}-1␣))␣1.00␣${1}␣120.0␣3␣180.00" >> ${2}
}

#
#Function to write stdin_file to evaluate energy
#
function stdin_file { #$1=job_base_name 2= fchk1 3=fchk2 4=natoms_fchk1 5=natoms_fchk2
    echo '{' > ${1}_stdin
    echo '␣basis_directory=␣"/basis_sets"' >> ${1}_stdin
    echo "␣name=␣${2}_${3}" >> ${1}_stdin
    echo '␣atom_groups=␣{' >> ${1}_stdin
    echo '␣␣keys=␣{␣name=␣atom_indices=␣fchk_file=␣rotation=␣shift=␣}' >> ${1}_stdin
    echo '␣␣data=␣{' >> ${1}_stdin
    echo $'␣␣␣Group1␣{␣1␣...␣'${4}$'␣}␣"'${2}$'"␣1␣0␣0␣0␣1␣0␣0␣0␣1␣0␣0␣0' >> ${1}_stdin
    echo $'␣␣␣Group2␣{␣'$(( ${4}+1 ))$'␣...␣'$(( ${4}+${5} ))$'␣}␣"'${3}$'"␣1␣0␣0␣0␣1␣0␣0␣0␣1␣0␣0␣0' >> ${1}_stdin
    echo '␣␣}' >> ${1}_stdin
    echo '␣}' >> ${1}_stdin
    echo '␣put_group_12_polarization_energy' >> ${1}_stdin
    echo '␣put_group_12_energies' >> ${1}_stdin
    echo '␣put_group_12_grimme2006_energy' >> ${1}_stdin
    echo '}' >> ${1}_stdin
}


#Functions to write PBS skripts for jobs
#
function gaussian_pbs { #$1=pbsname 2=ncpus 3=mem 4=comname 5=frame_name
    echo "#!/bin/sh" > $1
    echo "#PBS␣-V" >> $1
    echo "#PBS␣-l␣nodes=1:g09:ppn=${2}" >> $1
    echo "#PBS␣-j␣eo" >> $1
    echo "#PBS␣-l␣pmem=${3}gb" >> $1
```

```
    echo "#PBS␣-l␣walltime=1:00:00" >> $1
    echo "#PBS␣-q␣batch" >> $1
    echo "#PBS␣-p␣100" >> $1
    echo "#PBS␣-m␣a" >> $1
    echo "#PBS␣-M␣f.kleemiss@uni-bremen.de" >> $1
    echo "#PBS␣-N␣$4_$5" >> $1
    echo "FILENAME=$4" >> $1
    echo 'SERVER=$PBS_O_HOST' >> $1
    echo 'WORKDIR=/scratch/$USER/PBS_$PBS_JOBID' >> $1
    echo "SCP=/usr/bin/scp" >> $1
    echo "SSH=/usr/bin/ssh" >> $1
    echo 'SERVERPERMDIR=${SERVER}:$PBS_O_WORKDIR' >> $1
    echo 'mkdir␣/scratch/$USER/PBS_$PBS_JOBID' >> $1
    echo 'stagein()␣{' >> $1
    echo 'cd␣${WORKDIR}' >> $1
    echo '${SCP}␣-P␣2244␣${SERVERPERMDIR}/*.com␣$WORKDIR' >> $1
    echo '}' >> $1
    echo 'runprogram()␣{' >> $1
    echo 'g09␣${FILENAME}' >> $1
    echo '}' >> $1
    echo 'stageout()␣{' >> $1
    echo 'cd␣${WORKDIR}' >> $1
    echo "mv␣Test.FChk␣${4%.com}.FChk" >> $1
    echo '$SCP␣-P␣2244␣$WORKDIR/*.log␣$SERVERPERMDIR/' >> $1
    echo '$SCP␣-P␣2244␣$WORKDIR/*.FChk␣$SERVERPERMDIR/' >> $1
    echo 'rm␣-r␣$WORKDIR' >> ${1}
    echo '}' >> ${1}
    echo "stagein" >> $1
    echo "runprogram" >> $1
    echo "stageout" >> $1
    echo "exit" >> $1
}
function tonto_pbs { #$1=fragmentname $2=ligand $3=frame
    echo "#!/bin/sh" > ${1}_tonto.pbs
    echo "#PBS␣-V" >> ${1}_tonto.pbs
    echo "#PBS␣-l␣nodes=1:ppn=1" >> ${1}_tonto.pbs
    echo "#PBS␣-j␣eo" >> ${1}_tonto.pbs
    echo "#PBS␣-l␣pmem=30gb" >> ${1}_tonto.pbs
    echo "#PBS␣-l␣walltime=2:00:00" >> ${1}_tonto.pbs
    echo "#PBS␣-q␣batch" >> ${1}_tonto.pbs
    echo "#PBS␣-m␣a" >> ${1}_tonto.pbs
    echo "#PBS␣-M␣f.kleemiss@uni-bremen.de" >> ${1}_tonto.pbs
    echo "#PBS␣-N␣CE_$1_$2_$3" >> ${1}_tonto.pbs
    echo "FILENAME=$1" >> ${1}_tonto.pbs
    echo 'SERVER=$PBS_O_HOST' >> ${1}_tonto.pbs
    echo 'WORKDIR=/scratch/$USER/PBS_$PBS_JOBID' >> ${1}_tonto.pbs
    echo "SCP=/usr/bin/scp" >> ${1}_tonto.pbs
    echo "SSH=/usr/bin/ssh" >> ${1}_tonto.pbs
    echo 'SERVERPERMDIR=${SERVER}:$PBS_O_WORKDIR' >> ${1}_tonto.pbs
    echo 'mkdir␣/scratch/$USER/PBS_$PBS_JOBID' >> ${1}_tonto.pbs
    echo 'stagein()␣{' >> ${1}_tonto.pbs
    echo 'cd␣${WORKDIR}' >> ${1}_tonto.pbs
    echo 'sleep␣5' >> ${1}_tonto.pbs
    echo '${SCP}␣-P␣2244␣${SERVERPERMDIR}/*.FChk␣$WORKDIR' >> ${1}_tonto.pbs
    echo '${SCP}␣-P␣2244␣${SERVERPERMDIR}/*_stdin␣$WORKDIR' >> ${1}_tonto.pbs
    echo '${SCP}␣-P␣2244␣${SERVERPERMDIR}/../../tonto␣$WORKDIR' >> ${1}_tonto.pbs
    TEMP=$'${SCP}␣-P␣2244␣${SERVERPERMDIR}/../../'${2}$'.FChk␣$WORKDIR'
    echo $TEMP >> ${1}_tonto.pbs
    echo '}' >> ${1}_tonto.pbs
    echo 'runprogram()␣{' >> ${1}_tonto.pbs
    echo './tonto␣-i␣${FILENAME}_stdin␣-o␣${FILENAME}_stdout' >> ${1}_tonto.pbs
    echo 'rm␣./tonto' >> ${1}_tonto.pbs
    echo 'rm␣./*restricted' >> ${1}_tonto.pbs
    echo '}' >> ${1}_tonto.pbs
    echo 'stageout()␣{' >> ${1}_tonto.pbs
    echo 'cd␣${WORKDIR}' >> ${1}_tonto.pbs
    echo '$SCP␣-r␣-P␣2244␣$WORKDIR/*_stdout␣$SERVERPERMDIR/' >> ${1}_tonto.pbs
    echo 'rm␣-r␣$WORKDIR' >> ${1}
    echo '}' >> ${1}_tonto.pbs
    echo "stagein" >> ${1}_tonto.pbs
    echo "runprogram" >> ${1}_tonto.pbs
    echo "stageout" >> ${1}_tonto.pbs
    echo "exit" >> ${1}_tonto.pbs
}

#Run gaussian job for each frame on:
# 1) ligand
# 2) each fragment around the ligand in a seperate folder
# To do afterwards: Then merge all data into big folder above and run tonto job for each fragment
#
    gaussian_header ${ligand}.com
    echo "${charge}␣1" >> ${ligand}.com
    tail -n+3 ${ligand}.xyz >> ${ligand}.com
    echo "␣" >> ${ligand}.com
    gaussian_pbs ${ligand}.pbs $ncpus_g09 $mem_g09 ${ligand}.com main
    ligandjob=$(qsub ${ligand}.pbs)
    sleep 5

#
# separate large xyz trajectory into separate files:
#
    i=0
    for fragment in ${fragments[@]}; do
     NA=$(( ${natoms_fragments[$i]}+2 ))
     echo Splitting ${fragment}.xyz in smaller files with $NA lines
     i=$(( $i+1 ))
```

```
     split -a 4 -d -l ${NA} ${fragment}.xyz ${fragment}- --additional-suffix=.xyz
    done

    for frame in $(seq -f "%4g" 0 $(( ${frames}-1 ))); do
     printf -v j "%04d" ${frame}
     mkdir f_${j}
     cd f_${j}
     i=0
     for fragment in ${fragments[@]}; do
       mkdir ${fragment}
       cd $fragment
       mv ../../${fragment}-${j}.xyz ${fragment}.xyz

       gaussian_header ${fragment}.com
       echo "${charges[${i}]}␣1" >> ${fragment}.com
       tail -n+3 ${fragment}.xyz >> ${fragment}.com
       saturation ${natoms_fragments[${i}]} ${fragment}.com
       echo "␣" >> ${fragment}.com
       gaussian_pbs ${fragment}.pbs $ncpus_g09 $mem_g09 ${fragment}.com ${j}
       fragmentjob[${i}]=$(qsub ${fragment}.pbs)

       stdin_file $fragment ${ligand}.FChk ${fragment}.FChk $natoms_ligand ${natoms_fragments[${i}]}
       tonto_pbs $fragment $ligand ${j}
       tonto_job[${i}]=$(qsub –W depend=afterok:${ligandjob}:${fragmentjob[${i}]} ${fragment}_tonto.pbs)
       cd ..
       i=$(( $i+1 ))
     done
     cd ..
    done
```

Listing B.2: *bash* Script to gather Results form previous Script.

```
#!/bin/bash

#
#Give fragments used in calculation
#

    fragments=(arg glyala hsd leuser leutyr leu metval phe serleu trp tyr val1 val2)

#
#Give number of frames in trajectory files as argument to script
#

    frames=$1

    echo Exchange_repulsion > rep.ener
    echo Polarization > pol.ener
    echo Dispersion > disp.ener
    echo Electrostatic > ele.ener
    for fragment in ${fragments[@]}; do
        echo ${fragment} >> rep.ener
        echo ${fragment} >> pol.ener
        echo ${fragment} >> ele.ener
        echo ${fragment} >> disp.ener
    done
    for frame in $(seq -f "%4g" 0 $(( ${frames}-1 ))); do
        printf -v j "%04d" ${frame}
        echo ${j}
        cd f_${j}

            echo "␣" > rep.ener
            echo "␣" > pol.ener
            echo "␣" > disp.ener
            echo "␣" > ele.ener
            i=0

            for fragment in ${fragments[@]}; do
                cd $fragment

                    grep exch-rep\  ${fragment}_stdout | awk '{print␣$5}' >> ../rep.ener
                    grep Polarization ${fragment}_stdout | awk '{print␣$5}' >> ../pol.ener
                    grep E_coul\  ${fragment}_stdout | awk '{print␣$5}' >> ../ele.ener
                    grep Grimme06 ${fragment}_stdout | awk '{print␣$6}' >> ../disp.ener

                cd ..
                i=$(( $i+1 ))
            done

        cd ..

        paste -d"," ele.ener f_${j}/ele.ener >> ele_temp.ener
        paste -d"," rep.ener f_${j}/rep.ener >> rep_temp.ener
        paste -d"," pol.ener f_${j}/pol.ener >> pol_temp.ener
        paste -d"," disp.ener f_${j}/disp.ener >> disp_temp.ener
        mv ele_temp.ener ele.ener
        mv rep_temp.ener rep.ener
        mv pol_temp.ener pol.ener
        mv disp_temp.ener disp.ener

    done
```

Appendix C

Applying Quantum Crystallography for Complementary Bonding Analysis

C.1 Supporting Information for *Publication:* Chemical Bonding in Polarised Push-Pull Ethylenes

The measurements and refinements in chapters 3.1 were performed by me, including the first ever quantum crystallographic refinement of a twinned structure (4a), by integration of reflections only present in the major domain of the crystal. Figures S1 and S2 were produced by me, as well as tables S1, S2 and S3 and all of the text present in this chapter 3.1.
Also chapter 4.6 was prepared by me, where the idea of using the model compounds during the EDA was proposed and executed by me. All figures, tables and text were produced by me.

Supporting Information

Table of contents

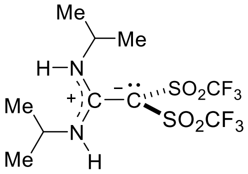
1. General and materials p. S1
2. Synthesis of push-pull ethylenes p. S2
3. X-ray crystallographic data p. S5
    3.1 X-ray crystallographic data of **3a**, **4a** and **5** with XWR p. S5
    3.2 X-ray crystallographic data of other push-pull ethylenes p. S9
4. Theoretical calculations p. S13
    4.1 Coordinates for optimized geometries of **3a** p. S13
    4.2 Coordinates for optimized geometries of **4a** p. S14
    4.3 Coordinates for optimized geometries of **5** p. S15
    4.4 DFT calculation for internal rotation around C1–C2 axis of **4b** p. S17
    4.5 NBO/NMR analysis p. S18
    4.6 Energy decomposition analysis p. S19
5. IR and Raman spectroscopy p. S21
6. ^{1}H and ^{13}C NMR spectra of all products p. S25
7. References p. S31

1. General and materials

All reactions were carried out under Ar atmosphere. Melting points were uncorrected. IR spectra were recorded on a Bruker ALPHA FT-IR spectrometer with an Attenuated Total Reflection (ATR) attachment. Raman spectra were recorded on a JASCO laser Raman spectrometer NRS-4100 (laser wavelength, 784 nm). NMR spectra were recorded on a Bruker Avance III Nanobay 400 MHz spectrometer (400 MHz for ^{1}H, 100 MHz for ^{13}C, and 376 MHz for ^{19}F) in $CDCl_3$ or CD_3CN. Data are reported as follows: chemical shifts, integration, multiplicity (s = singlet, d = doublet, t = triplet, q = quartet, quint = quintet, sex = sextet, m = multiplet, br = broad) and coupling constants. Chemical shifts (in ppm) in ^{1}H and ^{13}C NMR spectra were referenced to the solvent signal ($CDCl_3$, 7.26 ppm for ^{1}H NMR and 77.0 ppm for ^{13}C NMR; CD_3CN, 1.93 ppm for ^{1}H NMR and 117.7 ppm for ^{13}C NMR). Chemical shifts in ^{19}F NMR spectra were reported in ppm using trifluoromethylbenzene (0.00 ppm) as a standard. Coupling constants (J) are given in Hz. High resolution mass spectra (HRMS) were measured on a Micromass LCT mass spectrometer using electrospray ionization-time of flight (ESI-TOF). Column chromatography was performed on neutral silica gel (Kanto Chemical, Silica gel 60N, 63-210 μm). Tf_2CH_2 **2** was provided from Central Glass Co., Ltd. and this material can be also prepared by the Waller's procedure in the laboratory.[1] High-accuracy X-ray diffraction experiments were performed on a Bruker D8 Venture using Mo-Kα radiation from a microfocus source and a Photon100 CMOS detector in shutterless mode with open-flow nitrogen cooling.

2. Synthesis of push-pull ethylenes

N,N'-Diisopropyl-2,2-bis((trifluoromethyl)sulfonyl)ethene-1,1-diamine (3a)

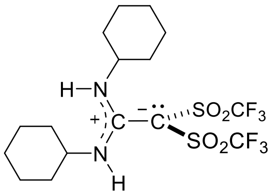


3a

According to the general procedure, this compound was isolated in 67% yield (135 mg, 0.332 mmol) by the reaction of Tf_2CH_2 **2** (139 mg, 0.496 mmol) and *N,N'*-diisopropylcarbodiimide (69.4 mg, 0.550 mmol) in CH_2Cl_2 (15 mL) for 3 h at room temperature and the following chromatographic purification (hexane/EtOAc = 5 : 1). The molecular structure was confirmed by a comparison of its NMR data with the reported one.[2]

IR (ATR) ν 3433, 3313, 1618, 1547, 1377, 1359, 1340, 1185, 1123, 1021, 617, 563, 503 cm^{-1}; ^1H NMR (400 MHz, CD_3CN) δ 1.20 (6H, d, J = 6.4 Hz), 1.29 (6H, d, J = 6.4 Hz), 3.71-3.85 (1H, m), 4.09-4.23 (1H, m), 7.01 (1H, brs), 7.48 (1H, brs); ^{13}C NMR (100 MHz, CD_3CN) δ 21.0, 22.7, 46.9, 51.4, 67.0, 121.6 (q, J_{CF} = 326 Hz), 154.8.

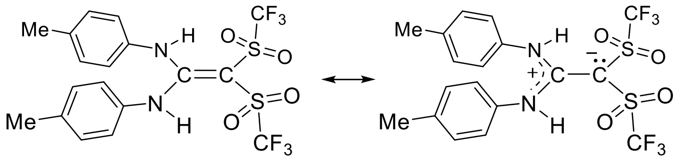
N,N'-Dicyclohexyl-2,2-bis((trifluoromethyl)sulfonyl)ethene-1,1-diamine (3b)



3b

According to the general procedure, this compound was isolated in 42% yield (98.5 mg, 0.202 mmol) by the reaction of Tf_2CH_2 **2** (136 mg, 0.485 mmol) with *N,N'*-dicyclohexylcarbodiimide (113 mg, 0.548 mmol) in CH_2Cl_2 (15 mL) for 6 h at room temperature and the following chromatographic purification (hexane/EtOAc = 5 : 1). Colorless crystals (from Et_2O); Mp. 190-192 °C; IR (ATR) ν 3321, 2936, 1615, 1365, 1192, 1151, 1024, 644, 634, 580, 501 cm^{-1}; ^1H NMR (400 MHz, CD_3CN) δ 1.09-1.52 (10H, m), 1.57-1.70 (2H, m), 1.70-2.08 (8H, m), 3.40-3.53 (1H, m), 3.67-3.81 (1H, m), 7.02 (1H, brs, N*H*), 7.41 (1H, brs, N*H*); ^{13}C NMR (100 MHz, CD_3CN) δ 25.1, 25.6, 25.8, 25.9, 31.5, 33.5, 53.6, 58.6, 67.3, 121.6 (q, J_{CF} = 326 Hz), 154.9; ^{19}F NMR (376 Hz, CD_3CN) δ –15.5 (6F, s); HRMS (ESI-TOF) calcd for $C_{16}H_{25}F_6N_2O_4S_2$ $[M+H]^+$, 487.1160; found, 487.1158; Anal. Calcd for $C_{16}H_{24}F_6N_2O_4S_2$: C, 39.50; H, 4.97; N, 5.76. Found: C, 39.33; H, 4.97; N, 5.76.

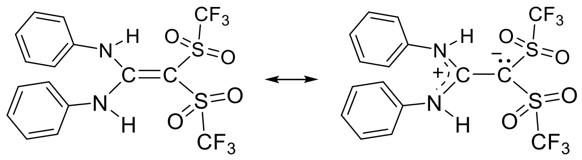
N,N'-Di-*p*-tolyl-2,2-bis((trifluoromethyl)sulfonyl)ethene-1,1-diamine (4a)



4a

According to the general procedure, this compound was isolated in 77% yield (185 mg, 0.368 mmol) by the reaction of Tf_2CH_2 **2** (134 mg, 0.478 mmol) with *N,N'*-di-*p*-tolylcarbodiimide (118 mg, 0.531 mmol) in CH_2Cl_2 (1.0 mL) for 3 h at room temperature and the following chromatographic purification (hexane/EtOAc = 5 : 1). Colorless crystals (from Et_2O); Mp. 214-215 °C; IR (ATR) ν 3312, 1594, 1523, 1513, 1374, 1205, 1187, 1116, 992, 806, 602, 576, 482 cm^{-1}; ^1H NMR (400 MHz, CD_3CN) δ 2.15 (6H, s), 6.80 (4H, d, J = 8.2 Hz), 6.92 (4H, d, J = 8.2 Hz), 9.35 (2H, brs, N*H*); ^{13}C NMR (100 MHz, CD_3CN) δ 20.8, 74,0, 121.5 (q, J_{CF} = 325 Hz), 123.3, 130.3, 133.9, 138.2, 153.5; ^{19}F NMR (376 Hz, CD_3CN) δ –16.5 (6F, s); MS (ESI-TOF) *m/z* 503 $[M+H]^+$; HRMS calcd for $C_{18}H_{17}F_6N_2O_4S_2$ $[M+H]^+$, 503.0534; found, 503.0518; Anal. Calcd for $C_{18}H_{16}F_6N_2O_4S_2$: C, 43.03; H, 3.21; N, 5.58. Found: C, 43.24; H, 3.40; N, 5.57.

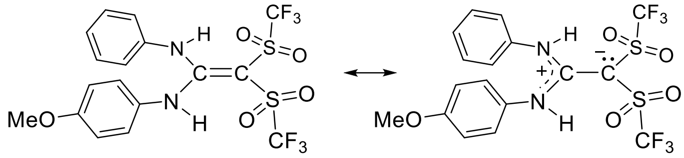
N,N'-Diphenyl-2,2-bis((trifluoromethyl)sulfonyl)ethene-1,1-diamine (4b)



4b

According to the general procedure, this compound was obtained in 69% yield (157 mg, 0.331 mmol) by the reaction of Tf_2CH_2 **2** (135 mg, 0.481 mmol) with *N,N'*-diphenlycarbodiimide[3] (103 mg, 0.530 mmol) in $ClCH_2CH_2Cl$ (15 mL) for 3 h at 50 °C and the following column chromatography on silica gel (hexane/EtOAc = 5 : 1). Colorless crystals (from $CHCl_3$); Mp. 209-210 °C; IR (ATR) ν 3416, 3312, 1574, 1539, 1371, 1192, 1120, 753, 645, 589, 495 cm^{-1}; ^1H NMR (400 MHz, CD_3CN) δ 6.93 (4H, brd, J = 7.6 Hz), 7.02 (2H, tt, J = 7.2, 2.0 Hz), 7.07-7.14 (4H, m), 9.47 (2H, brs, N*H*); ^{13}C NMR (100 MHz, CD_3CN) δ 74.3, 121.5 (q, J_{CF} = 325 Hz), 123.5, 128.0, 129.9, 136.4, 154.0; ^{19}F NMR (376 Hz, CD_3CN) δ –16.5 (6F, s); MS(ESI-TOF) *m/z* 497 $[M+Na]^+$; HRMS calcd for $C_{16}H_{12}F_6N_2NaO_4S_2$ $[M+Na]^+$, 497.0037; found, 497.0040; Anal. Calcd for $C_{16}H_{12}F_6N_2O_4S_2$: C, 40.51; H, 2.55; N, 5.91. Found: C, 40.61; H, 2.73; N, 5.99.

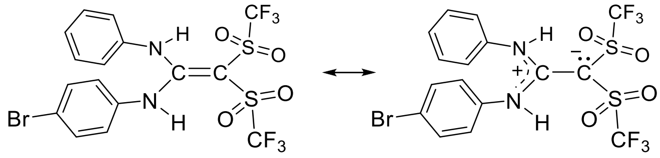
N-(4-Methoxyphenyl)-*N′*-phenyl-2,2-bis((trifluoromethyl)sulfonyl)ethene-1,1-diamine (4c)



4c

According to the synthetic procedure for **3a**, this compound was obtained in 95% yield (229 mg, 0.454 mmol) by the reaction of Tf_2CH_2 **2** (134 mg, 0.478 mmol) with *N*-(4-methoxyphenyl)-*N*'-phenylcarbodiimide[3] (124 mg, 0.553 mmol) in $ClCH_2CH_2Cl$ (15 mL), for 4 h at 50 °C and the following column chromatography on silica gel (hexane/EtOAc = 3 : 1). Yellow crystals (from $CHCl_3$); Mp. 173-175 °C; IR (ATR) ν 3334, 1575, 1367, 1196, 1123, 1010, 609, 575, 487 cm^{-1}; ^1H NMR (400 MHz, CD_3CN) δ 3.63 (3H, s), 6.62 (2H, d, J = 8.8 Hz), 6.85 (2H, d, J = 8.8 Hz), 6.90 (2H, d, J = 7.6 Hz), 7.04 (1H, t, J = 7.4 Hz), 7.08-7.16 (2H, m), 9.37 (2H, brs, N*H*); ^{13}C NMR (100 MHz, CD_3CN) δ 56.0, 74.0, 115.0, 121.5 (q, J_{CF} = 325 Hz), 123.7, 125.1, 127.9, 129.2, 129.7, 136.3, 153.8, 159.3; ^{19}F NMR (376 Hz, CD_3CN) δ –16.4 (6F, s); MS(ESI-TOF) *m/z* 527 $[M+Na]^+$; HRMS calcd for $C_{17}H_{14}F_6N_2NaO_5S_2$ $[M+Na]^+$, 527.0146; found, 527.0143; Anal. Calcd for $C_{17}H_{14}F_6N_2O_5S_2$: C, 40.48; H, 2.80; N, 5.55. Found: C, 40.77; H, 3.02; N, 5.61.

N-(4-Bromophenyl)-*N′*-phenyl-2,2-bis((trifluoromethyl)sulfonyl)ethene-1,1-diamine (4d)



4d

According to the general procedure, this compound was isolated in 70% yield (186 mg, 0.336 mmol) by the reaction of Tf_2CH_2 **2** (135 mg, 0.481 mmol) with *N*-(4-bromophenyl)-*N*'-phenylcarbodiimide[3] (151 mg, 0.553 mmol) in $ClCH_2CH_2Cl$ (15 mL) for 3 h at 50 °C and the following column chromatography on silica gel (hexane/EtOAc = 5 : 1). Colorless crystals (from $CHCl_3$); Mp. 170-173 °C; IR (ATR) ν 3312, 1590, 1376, 1206, 1115, 990, 651, 602, 502 cm^{-1}; ^1H NMR (400 MHz, CD_3CN) δ 6.84 (2H, d, J = 8.8 Hz), 6.93 (2H, d, J = 7.6 Hz), 7.05-7.11 (1H, m), 7.12-7.18 (2H, m), 7.25 (2H, d, J = 8.8 Hz), 9.43 (1H, brs, N*H*), 9.54 (1H, brs, N*H*); ^{13}C NMR (100 MHz, CD_3CN) δ 74.6, 120.9, 121.5 (q, J_{CF} = 325 Hz), 123.8, 125.5, 128.3 130.1, 132.9, 135.7, 136.3, 154.5; ^{19}F NMR (376 Hz, CD_3CN) δ –16.4 (6F, s); MS(ESI-TOF) *m/z* 575 $[M+Na]^+$; HRMS calcd for $C_{16}H_{11}BrF_6N_2NaO_4S_2$ $[M+Na]^+$, 574.9146; found, 574.9144; Anal. Calcd for $C_{16}H_{11}BrF_6N_2O_4S_2$: C, 34.73; H, 2.00; N, 5.06. Found: C, 34.68; H, 2.18; N, 5.14.

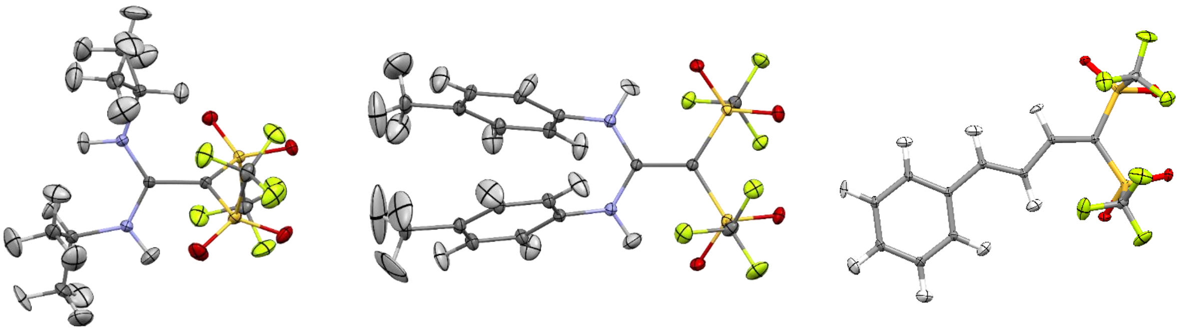
3. X-ray crystallographic data

Crystallographic data for the X-ray diffraction studies of **3a**, **3b**, **4a**-**4d** and **5** have been deposited with the Cambridge Crystallographic Data Center (CCDC) as supplementary publication nos. CCDC 1872536 (**3a**), 1873526 (**3b**) , 1872529 (**4a**), 1873527 (**4b**), 1873528 (**4c**), 1873525 (**4d**), 1872528 (**5**). These data can be obtained free of charge from the CCDC *via* www.ccdc.cam.ac.uk/data_request/cif.

3.1 X-ray crystallographic data of 3a, 4a and 5 with XWR

High quality and resolution X-ray diffraction experiments were performed for push-pull ethylenes **3a**, **4a**, and diene **5** at the in-house diffractometer (BRUKER D8-Venture in shutterless mode). The operation of the diffractometer was performed using the Software *APEX III* and the tools *SADABS*, *TWINABS*, *CELL_NOW* and *XPREP*. The crystallographic structures and information about the measurements and subsequent refinements are given in Figure S1 and Table S1, respectively. The structures were refined using Hirshfeld Atom Refinement (HAR) employing the programs *tonto*[4] and *Gaussian09*[5] interfaced through the self-written software *lamaGOET*. For HAR, the level of theory B3LYP/def2-TZVPP was used. Point charges and dipoles were imposed on every atom of a molecule that has at least one atom within a radius of 8 Å around the molecule of the asymmetric unit to simulate the crystal effect.

The structure of **4a** was solved after treating the twinning of the crystal used. There were two domains with mean I/σ of 17.0 in domain 1 and 6.2 in domain 2. R_{int} for domain 1 was 0.0291, for domain 2 it was 0.0603, so the overlapping reflections and the ones for domain 2 were disregarded, and only reflections of the first domain were taken into the refinement. For the sulfur atoms in all compounds, third and fourth order Gram-Charlier parameters were refined to treat effects of anharmonic atomic motions.



3a (CCDC 1872536) **4a** (CCDC 1872529) **5** (CCDC 1872528)

Figure S1. Structures of **3a**, **4a**, and **5** after XWR showing anisotropic displacement parameters for all atoms including hydrogens at a probability level of 50 %.

Table S1. Measurement, crystal structure and HAR refinement information for **3a**, **4a**, and **5**.

	3a	**4a**	**5**
Empirical formula	$C_{10}H_{16}N_2O_4F_6S_2$	$C_{18}H_{16}N_2O_4F_6S_2$	$C_{12}H_8O_4F_6S_2$
Formula weight (g mol^{-1})	406.37	502.46	394.30
Cell setting, space group	Monoclinic, $P2_1$/n	Triclinic, P-1	Monoclinic $P2_1$/n
Z	4	2	4
Temperature (K)	100		
Unit cell dimensions			
a (Å)	9.2809 (3)	8.660 (2)	10.0932 (2)
b (Å)	16.8556 (5)	10.116 (2)	12.9954 (3)
c (Å)	10.7429 (3)	12.390 (3)	11.8616 (3)
α (°)	90.0	88.50 (3)	90.0
β (°)	90.1060 (10)	82.78 (3)	110.1180 (10)
γ (°)	90.0	68.59 (3)	90.0
V (Å^3)	1680.56 (9)	1002.3 (4)	1460.90 (6)
Crystal size (mm)	0.178 x 0.213 x 0.394	0.112 x 0.254 x 0.302	0.146x0.271x0.338
Crystal Form, color	Block, colorless	Block, colorless	Block, yellow
Wavelength λ (Å)	0.71073		
Absorption Correction	Empirical multi-scan		
Absorption coefficient μ (mm^{-1})	0.397	0.350	0.451
Max. θ (°)	52.235	45.569	45.451
Completeness	100%	98.6%	100%
Redundancy	13.14	6.93	16.41
Weighting scheme	Based on measured s.u.'s[a]		
$R_{(sigma)}$	0.0084	0.0167	0.0131
Number of parameters	411	483	339
R_1(F)	0.0210	0.0218	0.0217
R_w(F)	0.0135	0.0174	0.0169
$\Delta\rho_{max}$, $\Delta\rho_{min}$ (e Å^{-3})	0.27, -0.32	0.32, -0.45	0.33, -0.46

[a] $w = 1 / \sigma$

After Hirshfeld Atom Refinement, the geometries were fixed and the molecular wavefunctions were fitted to the experimental structure factors. This full procedure is termed X-ray Wavefunction Refinement, see main text for more details and references. Here, the level of theory HF/def2-TZVPP without a cluster of point charges and dipoles was used because electron correlation and polarization via the crystal effect are to be extracted from the experimental data. The resulting effect is discussed in the main manuscript.

Based on these experimentally obtained XWR wavefunctions, a complementary bonding analysis was performed. The main results are reported in Tables 2 and 3 of the main manuscript. Here, we add the profile of the Laplacian of the electron density along the C1–C2 bond axis of the formal 'C=C' double bond of all three molecules (Figure S2) and a discussion of the hypervalency of the sulfur atoms in Tf groups according to Tables S2 and S3.

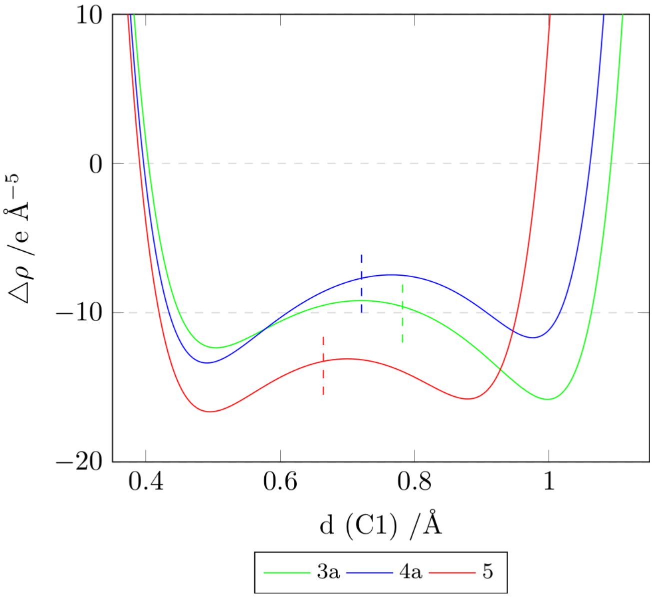


Figure S2. Laplacian of the electron density plotted along the C1–C2 bond vector (C1 on the left).

Table S2. XWR-derived atomic charge in *e*. For all compounds, the results confirm the resonance structure O^{-}–S^{2+}–O^{-}.

	3a		**4a**		**5**	
	QTAIM	NPA	QTAIM	NPA	QTAIM	NPA
S1	2.755	2.328	3.084	2.319	2.866	2.288
S2	2.801	2.324	2.930	2.319	2.711	2.293
O1	-1.390	-1.020	-1.430	-0.957	-1.399	-0.989
O2	-1.405	-0.967	-1.372	-1.021	-1.429	-0.953
O3	-1.385	-0.974	-1.428	-0.957	-1.276	-0.998
O4	-1.344	-1.022	-1.391	-1.021	-1.361	-0.956

Table S3. XWR-derived valency descriptors of the sufur atom.

		3a		**4a**		**5**	
		S1	S2	S1	S2	S1	S2
Σ NLMO/NPA BO[a]		3.309	3.340	3.296	3.295	3.288	3.272
Σ (ELI-D population) / *e* [b]		7.516	7.891	7.955	7.951	8.017	8.018
Σ DI [c]		3.893	3.883	3.743	3.825	3.778	3.864
Σ RGBI [d]		4.373	4.442	4.478	4.395	4.350	4.355
->Cov [e]		4.075	4.150	4.154	4.076	4.088	4.070
Durrant valency[f]	QTAIM	6.398	6.490	6.140	5.832	6.268	6.578
	NPA	7.344	7.352	7.362	7.362	7.424	7.414

[a] Sum of all four S–C and S–O NLMO/NPA bond orders from NRT analysis (hypervalent if greater than 4).

[b] Sum of all electron populations of the disynaptic S–C and S–O ELI-D bonding basins (hypervalent if greater than 8).

[c] Sum of all delocalisation indices of all four S–C and S–O bonds based on QTAIM (hypervalent if greater than 4).

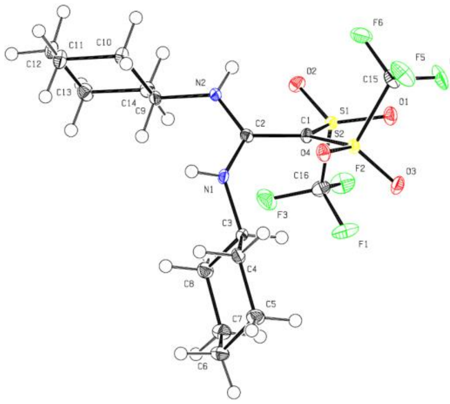
[d] Sum of the total RGBI of all four S-C and S–O bonds including covalent and ionic indices (not a measure of hypervalency).

[e] Sum of the covalent RGBI only (hypervalent if greater than 4).

[f] Durrant valency descriptors based on the QTAIM and NPA charges listed in Table S2 (hypervalent if greater than 8).[6]

3.2 X-ray crystallographic data of other push-pull ethylenes

Table S4. Crystallographic data of **3b** (CCDC 1873526)



$C_{16}H_{24}F_6N_2O_4S_2$
M_r = 486.49
Monoclinic, $P2_1/n$
Hall symbol: -P 2yn
a = 10.1959 (11) Å
b = 19.722 (2) Å
c = 10.3116 (12) Å
β = 92.117 (1)°
V = 2072.1 (4) Å^3
Z = 4

$F(000)$ = 1008
D_x = 1.559 Mg m^{-3}
Mo $K\alpha$ radiation, λ = 0.71073 Å
Cell parameters from 4702 reflections
θ = 2.2–27.6°
μ = 0.34 mm^{-1}
T = 90 K
Block, colourless
0.22 × 0.16 × 0.05 mm

Bruker APEXII CCD area detector diffractometer
Radiation source: Bruker TXS fine-focus rotating anode
Bruker Helios multilayer confocal mirror
Detector resolution: 8.333 pixels mm^{-1}
phi and ω scans
Absorption correction:
analytical Crystal Faces plugin in Bruker *APEX2* software
$T_{\min}$ = 0.930, $T_{\max}$ = 0.983
9987 measured reflections

3657 independent reflections
3269 reflections with $I > 2\sigma(I)$
R_{int} = 0.025
$\theta_{\max}$ = 25.0°, $\theta_{\min}$ = 2.1°
h = -12→12
k = -23→18
l = -12→12

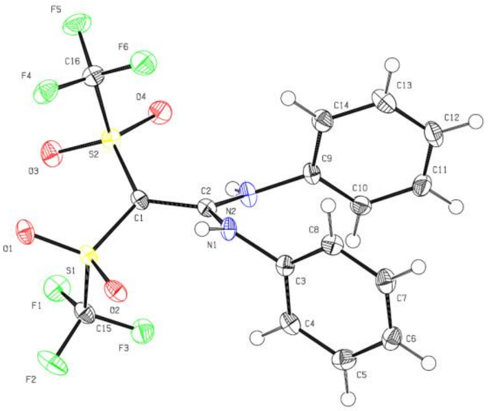
Refinement on F^2
Least-squares matrix: full
$R[F^2 > 2\sigma(F^2)]$ = 0.030
$wR(F^2)$ = 0.080
S = 1.03

Primary atom site location: structure-invariant direct methods
Secondary atom site location: difference Fourier map
Hydrogen site location: inferred from neighbouring sites
H atoms treated by a mixture of independent and constrained refinement
$w = 1/[\sigma^2(F_o^2) + (0.0395P)^2 + 1.2379P]$
where $P = (F_o^2 + 2F_c^2)/3$

3657 reflections
277 parameters
168 restraints

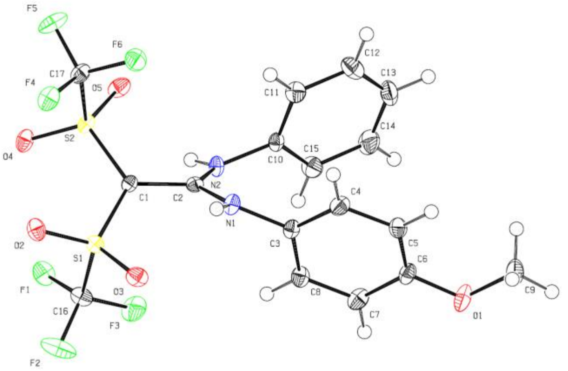
$(\Delta/\sigma)_{\max}$ = 0.001
$\Delta\rangle_{\max}$ = 0.64 e Å^{-3}
$\Delta\rangle_{\min}$ = -0.39 e Å^{-3}

Table S5. Crystallographic data of **4b** (CCDC 1873527)



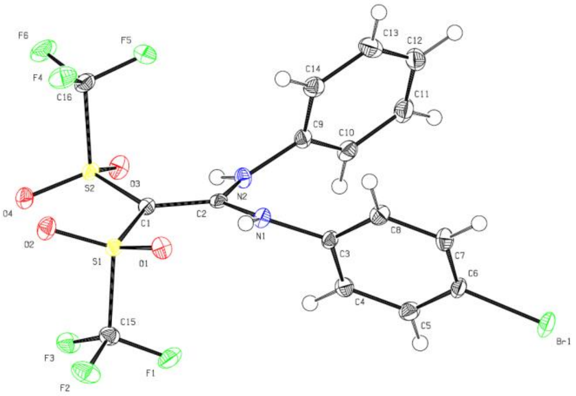
$C_{16}H_{12}F_6N_2O_4S_2$	$F(000) = 1920$
$M_r = 474.40$	$D_x = 1.678$ Mg m^{-3}
Monoclinic, $P2_1/c$	Mo $K\alpha$ radiation, $\lambda = 0.71073$ Å
Hall symbol: -P 2ybc	Cell parameters from 5987 reflections
$a = 15.8078$ (8) Å	$\theta = 2.6$–27.4°
$b = 21.4591$ (11) Å	$\mu = 0.37$ mm^{-1}
$c = 11.3453$ (6) Å	$T = 90$ K
$\beta = 102.607$ (1)°	Block, colourless
$V = 3755.8$ (3) Å^3	0.13 × 0.07 × 0.03 mm
$Z = 8$	
Bruker APEXII CCD area detector diffractometer	6618 independent reflections
Radiation source: Bruker TXS fine-focus rotating anode	5549 reflections with $I > 2\sigma(I)$
Bruker Helios multilayer confocal mirror	$R_{\mathrm{int}} = 0.028$
Detector resolution: 8.333 pixels mm^{-1}	$\theta_{\max} = 25.0°$, $\theta_{\min} = 1.3°$
phi and ω scans	$h = -18 \rightarrow 18$
Absorption correction:	$k = -25 \rightarrow 20$
analytical Crystal Faces plugin in Bruker *APEX2* software	
$T_{\min} = 0.954$, $T_{\max} = 0.989$	$l = -10 \rightarrow 13$
18201 measured reflections	
Refinement on F^2	Primary atom site location: structure-invariant direct methods
Least-squares matrix: full	Secondary atom site location: difference Fourier map
$R[F^2 > 2\sigma(F^2)] = 0.031$	Hydrogen site location: inferred from neighbouring sites
$wR(F^2) = 0.078$	H atoms treated by a mixture of independent and constrained refinement
$S = 1.04$	$w = 1/[\sigma^2(F_o^2) + (0.0347P)^2 + 1.740P]$
	where $P = (F_o^2 + 2F_c^2)/3$
6618 reflections	$(\Delta/\sigma)_{\max} = 0.001$
553 parameters	$\Delta\rangle_{\max} = 0.49$ e Å^{-3}
324 restraints	$\Delta\rangle_{\min} = -0.38$ e Å^{-3}

Table S6. Crystallographic data of **4c** (CCDC 1873528)



$C_{17}H_{14}F_6N_2O_5S_2$	$Z = 2$
$M_r = 504.42$	$F(000) = 512$
Triclinic, $P\bar{1}$	$D_x = 1.664$ Mg m^{-3}
Hall symbol: -P 1	Mo $K\alpha$ radiation, $\lambda = 0.71073$ Å
$a = 8.7659$ (5) Å	Cell parameters from 5751 reflections
$b = 10.8741$ (6) Å	$\theta = 2.5–27.6°$
$c = 11.4322$ (6) Å	$\mu = 0.35$ mm^{-1}
$\alpha = 96.618$ (1)°	$T = 90$ K
$\beta = 92.241$ (1)°	Block, colourless
$\gamma = 110.968$ (1)°	0.09 × 0.08 × 0.05 mm
$V = 1006.99$ (10) Å^3	
Bruker APEXII CCD area detector diffractometer	3562 independent reflections
Radiation source: Bruker TXS fine-focus rotating anode	3239 reflections with $I > 2\sigma(I)$
Bruker Helios multilayer confocal mirror	$R_{int} = 0.017$
Detector resolution: 8.333 pixels mm^{-1}	$\theta_{max} = 25.0°$, $\theta_{min} = 1.8°$
phi and ω scans	$h = -10 \rightarrow 10$
Absorption correction:	$k = -12 \rightarrow 12$
analytical Crystal Faces plugin in Bruker *APEX2* software	
$T_{min} = 0.969$, $T_{max} = 0.983$	$l = -13 \rightarrow 13$
9970 measured reflections	
Refinement on F^2	Primary atom site location: structure-invariant direct methods
Least-squares matrix: full	Secondary atom site location: difference Fourier map
$R[F^2 > 2\sigma(F^2)] = 0.026$	Hydrogen site location: inferred from neighbouring sites
$wR(F^2) = 0.067$	H atoms treated by a mixture of independent and constrained refinement
$S = 1.03$	$w = 1/[\sigma^2(F_o^2) + (0.0312P)^2 + 0.611P]$ where $P = (F_o^2 + 2F_c^2)/3$
3562 reflections	$(\Delta/\sigma)_{max} = 0.001$
296 parameters	$\Delta\rangle_{max} = 0.43$ e Å^{-3}
174 restraints	$\Delta\rangle_{min} = -0.34$ e Å^{-3}

Table S7. Crystallographic data of **4d** (CCDC 1873525)



$2(C_{16}H_{11}BrF_6N_2O_4S_2)\cdot CHCl_3$	V = 1103.50 (9) Å^3
M_r = 1213.95	Z = 1
Triclinic, $P\bar{1}$	$F(000)$ = 600
a = 8.6610 (4) Å	D_x = 1.827 Mg m^{-3}
b = 10.3611 (5) Å	Mo $K\alpha$ radiation, λ = 0.71073 Å
c = 13.7308 (7) Å	μ = 2.31 mm^{-1}
α = 95.043 (1)°	T = 90 K
β = 103.089 (1)°	Block, colourless
γ = 110.693 (1)°	0.09 × 0.08 × 0.08 mm
Bruker APEXII CCD area detector diffractometer	3889 independent reflections
Radiation source: Bruker TXS fine-focus rotating anode	3638 reflections with $I > 2\sigma$
Bruker Helios multilayer confocal mirror monochromator	R_{int} = 0.017
Detector resolution: 8.333 pixels mm^{-1}	θ_{max} = 25.0°, θ_{min} = 1.6°
phi and ω scans	h = -10→10
Absorption correction:	k = -12→12
numerical Crystal Faces plugin in Bruker *APEX2* software	
T_{min} = 0.782, T_{max} = 0.837	l = -16→16
10877 measured reflections	
Refinement on F^2	216 restraints
Least-squares matrix: full	Hydrogen site location: mixed
$R[F^2 > 2\sigma(F^2)]$ = 0.020	H atoms treated by a mixture of independent and constrained refinement
$wR(F^2)$ = 0.049	$w = 1/[\sigma^2(F_o^2) + (0.0232P)^2 + 0.626P]$ where $P = (F_o^2 + 2F_c^2)/3$
S = 1.03	$(\Delta/\sigma)_{max}$ = 0.001
3889 reflections	$\Delta\rangle_{max}$ = 0.39 e Å^{-3}
324 parameters	$\Delta\rangle_{min}$ = -0.32 e Å^{-3}

4. Theoretical calculations

Experimentally established geometries of **3a**, **4a**, and **5** were optimized and characterized by frequency analysis using hybrid density functional theory (M06-2x)[7] and the 6-311+G(d,p) basis set as implemented in the *Gaussian 09* program, revision D.01.[5] Key interatomic distances are summarized in Table S8.

Table S8. Interatomic distances of **3a**, **4a**, and **5**

		Interatomic distance (pm)						
		C1–C2	C1–N[a]		C2–S[b]		S–CF_3[c]	
3a	Exp.	147.98	132.76	132.20	169.81	168.23	184.87	185.23
	Calc.	145.2	133.4	132.9	171.9	172.5	186.1	186.8
4a	Exp.	145.64	133.98	134.00	172.09	171.92	185.21	185.36
	Calc.	143.9	134.4	134.4	174.4	174.4	186.4	185.4
5	Exp.	137.45	110.21[d]	142.22[e]	175.87	174.93	185.33	186.26
	Calc.	135.7	108.9[d]	143.3[e]	178.8	177.6	187.1	187.1

[a] C1–N1 and C1–N2. [b] C1–S1 and C1–S2. [c] S1–C3 and S2–C4. [d] C1–H1, [e] C1–C5

4.1 Coordinates for optimized geometries of 3a

```
---------------------------------------------------------------------
 Center     Atomic      Atomic             Coordinates (Angstroms)
 Number     Number       Type             X           Y           Z
---------------------------------------------------------------------
      1          6           0       -0.714465   -0.909058   -0.311630
      2          6           0        0.323636    0.069429   -0.042924
      3          6           0       -1.069942   -1.631586    2.088828
      4          1           0       -0.310774   -0.882076    2.303221
      5          6           0       -0.645842   -2.979120    2.657489
      6          1           0       -1.410100   -3.736431    2.452608
      7          1           0       -0.540489   -2.901673    3.740884
      8          1           0        0.304299   -3.294280    2.231690
      9          6           0       -2.414528   -1.175627    2.651129
     10          1           0       -2.698677   -0.200315    2.255620
     11          1           0       -2.352836   -1.104592    3.738724
     12          1           0       -3.199485   -1.899238    2.408185
     13          6           0       -2.554960   -1.493268   -1.900802
     14          1           0       -2.616906   -2.541358   -1.586966
     15          6           0       -2.663938   -1.447451   -3.419331
     16          1           0       -2.590774   -0.412981   -3.766962
     17          1           0       -3.627910   -1.844764   -3.737853
     18          1           0       -1.872266   -2.033822   -3.888528
     19          6           0       -3.651754   -0.667863   -1.228110
     20          1           0       -3.574034   -0.695961   -0.139558
     21          1           0       -4.633910   -1.049570   -1.513397
     22          1           0       -3.569310    0.373421   -1.547951
     23          6           0        2.941697   -0.608210   -0.860565
     24          6           0       -0.973719    2.387108    0.827144
     25          9           0        2.338328   -1.398447   -1.748942
     26          9           0        4.076613   -1.173130   -0.481838
     27          9           0        3.195728    0.554517   -1.427571
     28          9           0       -2.108334    1.678027    0.886650
     29          9           0       -1.273159    3.650109    0.583060
```

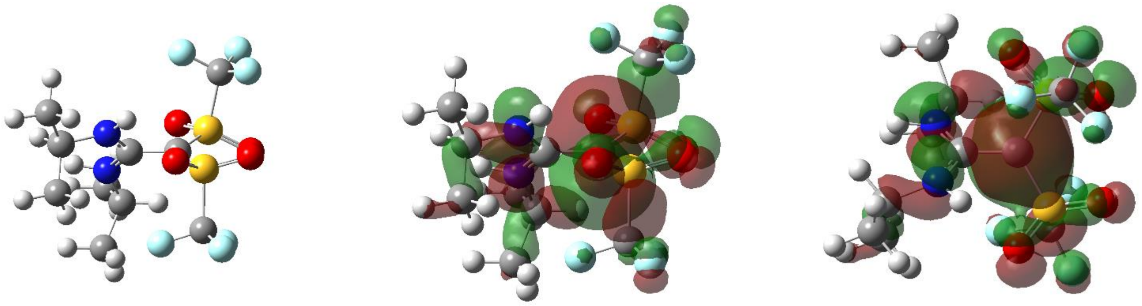
30	9	0	-0.359941	2.293153	1.993937
31	7	0	-1.177619	-1.733638	0.621420
32	7	0	-1.225813	-0.977703	-1.541585
33	8	0	1.684913	-1.797696	1.085049
34	8	0	2.443591	0.579114	1.467158
35	8	0	-0.800513	1.749922	-1.692143
36	8	0	1.312390	2.455660	-0.502453
37	16	0	1.832891	-0.424077	0.631731
38	16	0	0.087768	1.698615	-0.537162
39	1	0	-1.796671	-2.468251	0.301156
40	1	0	-0.819965	-0.321522	-2.198755

E(RM062X) = -2196.28315155

Zero-point correction= 0.289725 (Hartree/Particle)

Sum of electronic and thermal Enthalpies= -2195.966349

Sum of electronic and thermal Free Energies= -2196.050673



(A) (B) (C)

Figure S3. Optimized structure of **3a** and its HOMO. (A) Optimized structure from side view, (B) HOMO from side view, (C) HOMO from top view. MOs are vizualized by *GaussView 5.0* (isovalue = 0.02).

4.2 Coordinates for optimized geometries of 4a

Center Number	Atomic Number	Atomic Type	Coordinates (Angstroms) X	Y	Z
1	6	0	1.617852	0.000002	0.000010
2	6	0	0.179091	0.000017	0.000013
3	6	0	-1.823550	-1.255527	0.647070
4	6	0	-2.032724	-2.095788	-0.439783
5	1	0	-1.201058	-2.351635	-1.086873
6	6	0	-3.308896	-2.582364	-0.692129
7	1	0	-3.472718	-3.231370	-1.545948
8	6	0	-4.382900	-2.251687	0.135758
9	6	0	-4.142700	-1.427671	1.236422
10	1	0	-4.964481	-1.158408	1.891381
11	6	0	-2.872681	-0.930473	1.496749
12	1	0	-2.695468	-0.263887	2.332431
13	6	0	-5.761041	-2.796004	-0.133680
14	1	0	-5.896365	-3.761930	0.360757
15	1	0	-5.924596	-2.945482	-1.201937
16	1	0	-6.530845	-2.120186	0.241783
17	6	0	-1.823544	1.255565	-0.647054
18	6	0	-2.872688	0.930426	-1.496682

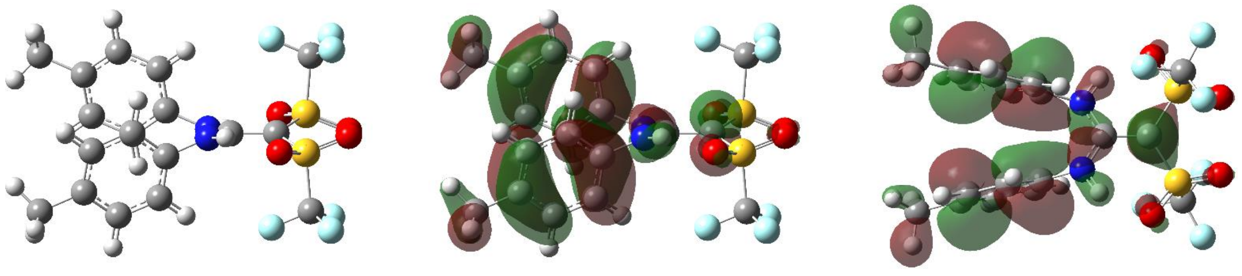
19	1	0	-2.695481	0.263779	-2.332316
20	6	0	-4.142714	1.427607	-1.236353
21	1	0	-4.964509	1.158269	-1.891262
22	6	0	-4.382905	2.251693	-0.135740
23	6	0	-3.308889	2.582445	0.692104
24	1	0	-3.472711	3.231487	1.545896
25	6	0	-2.032712	2.095890	0.439752
26	1	0	-1.201038	2.351786	1.086812
27	6	0	-5.761039	2.796063	0.133618
28	1	0	-5.924685	2.945495	1.201865
29	1	0	-6.530846	2.120315	-0.241962
30	1	0	-5.896256	3.762025	-0.360784
31	6	0	2.692116	-2.565343	-0.477672
32	6	0	2.692198	2.565321	0.477660
33	9	0	1.467508	-2.934225	-0.857579
34	9	0	3.319895	-3.608562	0.033491
35	9	0	3.349788	-2.117330	-1.528811
36	9	0	1.467607	2.934238	0.857584
37	9	0	3.319998	3.608518	-0.033520
38	9	0	3.349875	2.117296	1.528791
39	7	0	-0.512771	-0.724159	0.895759
40	7	0	-0.512759	0.724207	-0.895730
41	8	0	1.717277	-1.858804	1.847085
42	8	0	3.877040	-0.802210	1.126243
43	8	0	1.717306	1.858792	-1.847079
44	8	0	3.877049	0.802143	-1.126258
45	16	0	2.540612	-1.233790	0.818205
46	16	0	2.540637	1.233762	-0.818205
47	1	0	0.037352	-1.183168	1.617304
48	1	0	0.037367	1.183207	-1.617278

E(RM062X) = -2501.09654106

Zero-point correction= 0.337191 (Hartree/Particle)

Sum of electronic and thermal Enthalpies= -2500.728235

Sum of electronic and thermal Free Energies= -2500.822986



(A) (B) (C)

Figure S4. Optimized structure of **4a** and its HOMO. (A) Optimized structure from side view, (B) HOMO from sid view, (C) HOMO from top view. MOs are vizualized by *GaussView 5.0* (isovalue = 0.02).

4.3 Coordinates for optimized geometries of 5

Center Number	Atomic Number	Atomic Type	Coordinates (Angstroms) X	Y	Z
1	6	0	-0.876690	-0.355187	-0.589023
2	6	0	0.315700	-0.949888	-0.331298

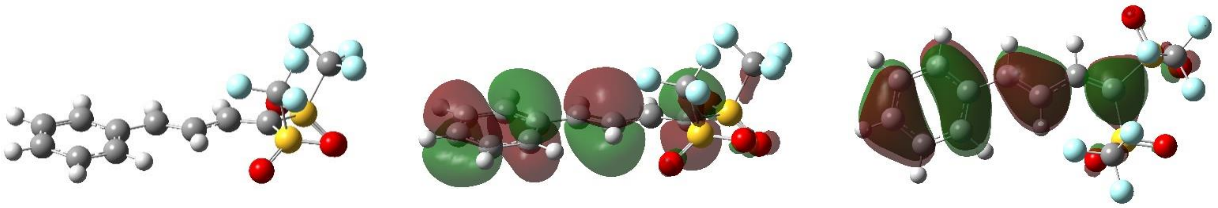
3	1	0	0.253696	-2.015470	-0.113482
4	6	0	1.631230	-0.381725	-0.329044
5	1	0	1.763404	0.656784	-0.596248
6	6	0	2.678588	-1.172398	-0.007281
7	1	0	2.474249	-2.209703	0.252743
8	6	0	4.080673	-0.779823	0.039821
9	6	0	4.522881	0.490766	-0.358444
10	1	0	3.817462	1.222323	-0.733905
11	6	0	5.867169	0.815922	-0.288104
12	1	0	6.200081	1.797434	-0.602406
13	6	0	6.791672	-0.116917	0.181264
14	1	0	7.842197	0.142974	0.234351
15	6	0	6.366932	-1.380873	0.575875
16	1	0	7.082748	-2.107856	0.939514
17	6	0	5.020438	-1.709972	0.501953
18	1	0	4.685911	-2.695210	0.808595
19	6	0	-0.811490	2.108535	0.769705
20	6	0	-3.325447	-1.036370	0.805315
21	9	0	-1.674908	1.569652	1.614365
22	9	0	-1.009739	3.410036	0.695329
23	9	0	0.421039	1.871720	1.202666
24	9	0	-3.848011	0.169851	0.729423
25	9	0	-4.288636	-1.936966	0.875116
26	9	0	-2.558694	-1.122664	1.883627
27	8	0	-2.393330	1.689722	-1.316076
28	8	0	0.107892	1.829833	-1.691690
29	8	0	-3.096543	-1.097652	-1.859769
30	8	0	-1.818742	-2.774047	-0.464765
31	16	0	-2.300167	-1.429711	-0.709120
32	16	0	-1.040668	1.378388	-0.938376

E(RM062X) = -2158.12679912

Zero-point correction= 0.199728 (Hartree/Particle)

Sum of electronic and thermal Enthalpies= -2157.903714

Sum of electronic and thermal Free Energies= -2157.981923



(A) (B) (C)

Figure S5. Optimized structure of **3c** and its HOMO. (A) Optimized structure from side view, (B) HOMO from side view, (C) HOMO from top view. MOs are vizualized by GaussView 5.0 (isovalue = 0.02).

4.4 DFT calculation for internal rotation around the C1–C2 axis of 4b

To obtain quantitative information on internal rotation around the ‘C=C’ partial bond, we calculated the rotational profile of C_2-symmetric diphenyl derivative **4b** as a model by *Firefly 8.0* (Figure S6).[8] By rotating a N1–C2–N2 plane of stable conformer **I** (τ = 15.2°) in an anticlockwise manner, we found the first transition state **TS1**, the sub-stable conformer **II**, and the second transition state **TS2**. The difference of the C1–C2 interatomic distance was not changed significantly during the rotation (145.5±2.2 pm). With a small rotational barrier $\Delta H_{\mathrm{corr.}}$ (6.5 kcal mol^{-1}), easy rotation around the bond in the solution phase can be assumed. Moreover, these data suggest that intramolecular hydrogen bonding between the N–H part and the sulfonic oxygen can keep the C1–C2 geometry nearly planar, whereas the steric bulkiness of the hydrogen atoms works as a separator for the interatomic distance.

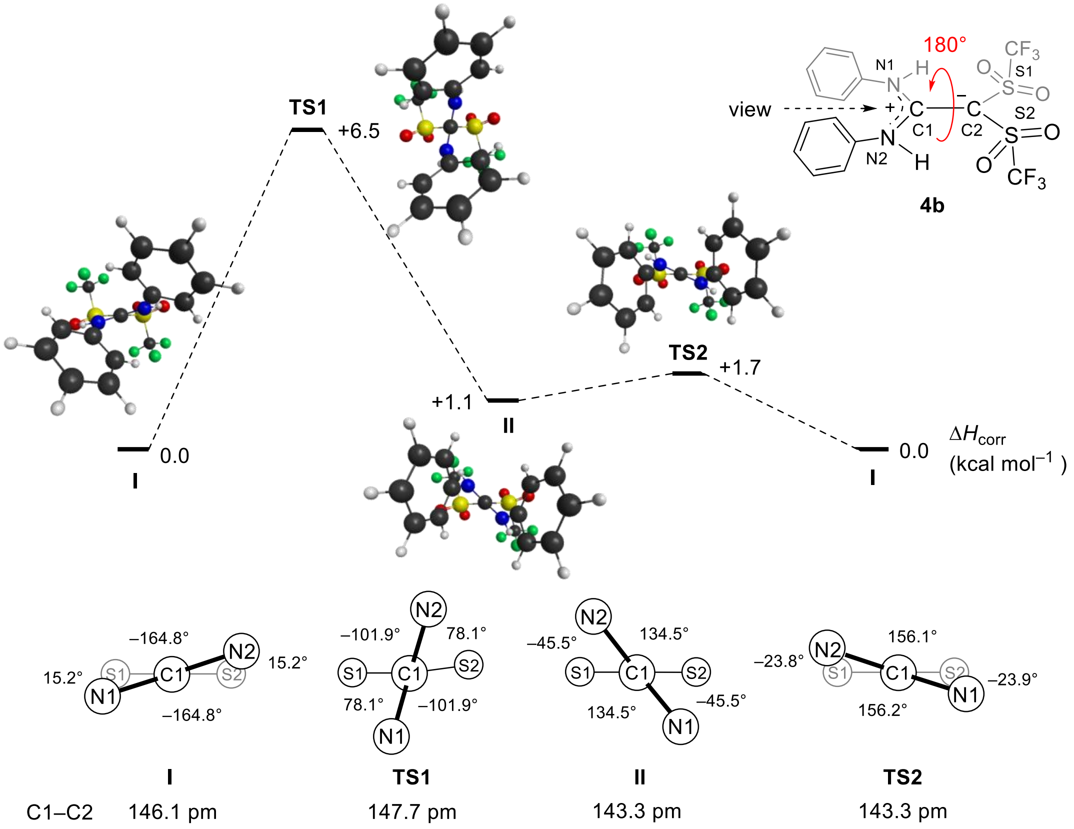
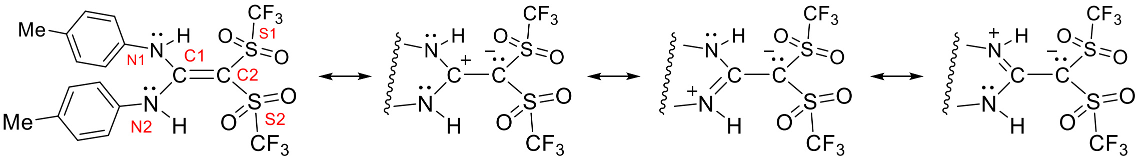


Figure S6. Reaction profile for internal rotation of **4b** around the C1–C2 axis [B3LYP/6-311+G(d)].

4.5 NBO/NRT analysis

With the DFT optimized geometry of **4a**, an NBO/NRT analysis was conducted by using the *NBO 6.0* program (Table 9).[9] Here in order to evaluate the effects of initial wavefuctions, we examined HF/6-311+G(d,p), HF/def2-TZVPP, B3LYP/6-311+G(d,p), M06-2x/6-311+G(d,p) calculations. In the NBO/NRT analysis using DFT wavefunctions, we found the π bonding resonance structure **A** as the major contributor. In contrast, the analysis of HF wavefunctions, even by using the def2-TZVPP basis set, yielded the charge-separated resonance structure **B1** as the major one. By using HF/def2-TZVPP level of theory, we also examined a difference between 'global' NRT data and 'local' NRT data for a N1(N2)C1C2(S1)S2 subset.

Table S9. NBO/NRT data **4a** using theoretical wavefunction



4a **A** **B1** **B2** **B3**

		HF 6-311+G(d,p)	HF def2-TZVPP	B3LYP 6-311+G(d,p)	M06-2x 6-311+G(d,p)
The most contributed structure		**B1**	**B1**	**A**	**A**
NPA charge	C1	0.680	0.694	0.492	0.524
	C2	–0.913	–1.022	–0.818	–0.848
Occupacy (*e*)	LV_{C1}	0.703	0.698	--	--
	LP_{C2}	1.658	1.650	--	--
	$\pi_{C1–C2}$	--	--	1.846	1.855
	$\pi^*_{C1–C2}$	--	--	0.574	0.562
E(2) (kcal mol^{-1})	LP_{C2} / LV_{C1}	209.7	208.4	--	--
	$LP_{C2} / \sigma^*_{S-CF3}$	17.6, 17.6	20.2, 20.2	--	--
	$\pi_{C1–C2} / \pi^*_{S–CF3}$	--	--	8.4, 8.4	10.1, 10.1
	LP_N / LV_{C2}	190.2, 190.2	188.5, 188.5	--	--
	$LP_N / \pi^*_{C1–C2}$	--	--	82.0, 82.1	105.6, 105.6
NLMO/NPA bond order		1.149	1.134	1.251	1.239
NRT bond order[a]	*Total*	1.147	1.159 (1.108)	1.676	1.682
	Covalent	1.003	0.987 (0.993)	1.269	1.252
	Ionic	0.144	0.172 (0.116)	0.407	0.430
Resonance weight[a,b] (**A**/**B**, %)		15/85	17/82 (11/89)	65/33	66/32

[a] Values in a 'local' subset, N1(N2)C1C2(S1)S2. Values of 'global' NRT analysis are shown in parentheses.

[b] Sum of contribution wieights by the resonance strcutures. Due to exsitance of resonance structures, which cannot be represented by **A**, **B1**, **B2**, or **B3**, total percentages did not always add up to 100%.

4.6 Energy decomposition analysis

An energy decomposition analysis (EDA)[30] allows evaluation of Pauli-repulsion, electrostatic and orbital contributions to the overall interaction energies between two fragments according to

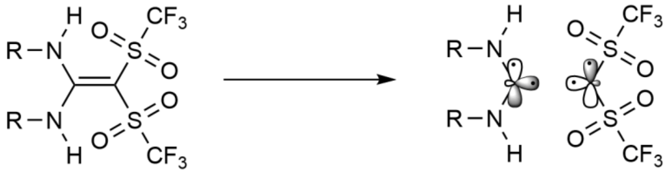
$$\Delta E_{Int} = E_{Pauli} + E_{Ele} + E_{Orb}.$$

EDA of the double bond in the compounds of question using M06-2X/TZVP in ADF (2017.114)[10,11] resulted in the following values in kJ/mol:

	E_{Pauli}	E_{Ele}	E_{Orb}	ΔE_{Int}
3a	1375.1	-856.6	-1642.1	-1123.6
4a	1549.6	-955.6	-1661.3	-1067.3
5	1509.5	-894.7	-1471.7	-856.9
3a(N)	1139.0	-698.1	-1072.1	-631.2
4a(N)	1112.5	-698.8	-1232.1	-818.5
5(N)	1366.2	-817.3	-1389.3	-840.4
3a(S)	1125.3	-756.7	-1173.0	-804.5
4a(S)	1214.8	-804.4	-1256.8	-846.3
5(S)	2497.7	-793.7	-2483.6	-779.7

Here the names are according to the main paper. Geometries of the two interacting fragments were not optimized, but taken from the experiment directly to be consistent with the XWR analyses. (**N**) denotes a model system, where the substituents of the NH*p*-Tol/NH*i*-Pr or CH=CHPh groups were exchanged by a single hydrogen atom, keeping angles between them fixed and changing only the distance of the hydrogen atom to 1.115 Å. This was done, because the interaction of fragments of this large size is not only due to the covalent C–C bond in question, but also considers electrostatics of side chains or hydrogen bonds broken by separating the fragments. Accordingly, (**S**) denotes the model compound where each of the Tf groups is replaced by a single H atom.

The orbital choice for the fragments was done according to the following scheme:



Biradical fragments were chosen, where the two SOMOs were of a shape that resembles one sp^n-hybrid orbital and one *p*-type orbital axial to the bond which was broken. Identical selection was carried out for the model systems.

The comparison of **3a**/**4a** to **5** is somewhat misleading if we want to focus on the C=C double-bond character, as the secondary interactions are in the same order of magnitude than direct influences of the C–C bonding, which is revealed by comparing each of them to the corresponding **(N)** systems. Comparison of **3a** and **4a** shows that Pauli repulsion drastically increases when “untwisting” the bond by 174.5 kJ/mol. Meanwhile electrostatic interactions also increase by –99.0 kJ/mol. Orbital energy is increased by –19 kJ/mol, which in total leads to a decrease of the total bonding energy of 56.4 kJ/mol for the planar structure compared to the twisted structure.

But considering the presence of hydrogen bonds and other interfragment effects, the trends in the model systems seem much clearer. The Pauli repulsion and electrostatic terms for **3a(N)** and **4a(N)** are almost identical, while the orbital overlap drastically increases by 160 kJ/mol when going from **3a** to **4a**. **5(N)** shows an even stronger Pauli repulsion, but also much stronger electrostatic and orbital interaction, which is most likely due to the 0.1 Å shorter C1–C2 bond compared to **3a** and **4a**. The model compounds show quite directly how the C1–C2 bond is affected by the different geometries and interactions with the Tf groups, without being interfered by secondary interactions.

Assuming the total (T) interfragment interactions in the systems can be broken down into partial energies, the following scheme was used to merge model energies of the whole system into a partial sum between functional groups, where a “/” between functional groups denotes interaction:

$$E_T = E_{C_1-C_2} + E_{SO_2-CF_3/C_1-C_2} + E_{NH-R/C_1-C_2} + E_{SO_2-CF_3/NH-R}$$

$$E_{(N)} = E_{C_1-C_2} + E_{SO_2-CF_3/C_1-C_2}$$

$$E_{(S)} = E_{C_1-C_2} + E_{NH-R/C_1-C_2}$$

$$E_T - (E_{(S)} + E_{(N)}) = E_{SO_2-CF_3/NH-R} - E_{C_1-C_2}$$

In this assumption, the resulting value (in kJ/mol) would correspond to the difference in energy of secondary interactions of fragments and the C–C bond.

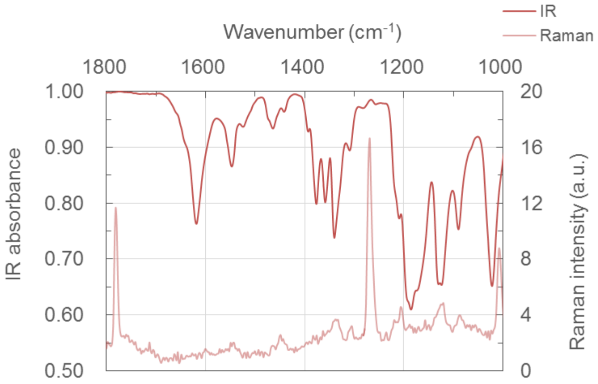
	E_T–(E_S+E_N)			
	Pauli	Ele	Orb	Int
3a	-889.2	598.2	603.0	312.0
4a	-777.7	547.7	827.7	597.6
5	-2354.4	716.4	2401.2	763.2

If we assume that both interactions are attractive in general, meaning having negative values, positive values show that the total bonding energy C1–C2 bond is much more significant than the secondary interactions between the two substituent groups. In total, a trend can be observed that the interaction going from **3a** over **4a** to **5** increases the interaction energy drastically. This is in good agreement with the observed bond properties in other techniques.

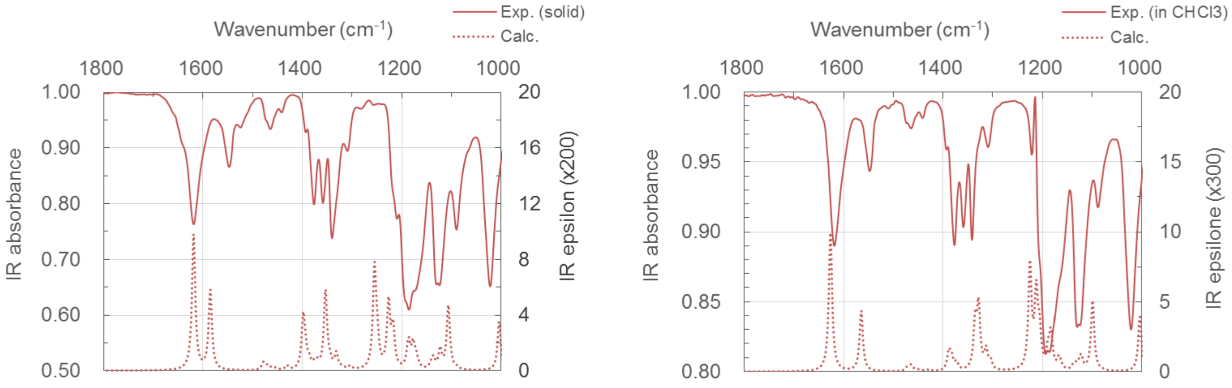
5. IR and Raman spectroscopy

To obtain vibrational information on formal 'C=C' bonds in the push-pull ethylenes, we measured IR and Raman spectra of **3a**, **4a**, and **5** in the solid state (Figures S7A, S9A, and S11A). IR spectra of them were also measured by using solutions in $CHCl_3$ (5.0 mmol L^{-1} for **3a** and **4a**, and 10 mmol L^{-1} for **5**). The solid state IR data are compared with IR spectra obtained by frequency calculations using the DFT-optimised geometry (see, chapter 4) (Figures S7B, S9B, and S11B). The IR data in $CHCl_3$ solutions are compared with IR spectra obtained by frequency calculations using the DFT-optimised geometry at PCM($CHCl_3$)-M06-2x/6-311+G(d,p) level of theory (Figures S7C, S9C, and S11C). Selected vibration vectors in the free molecules are shown in Figure S8, S10, and S12.

In IR spectrum of diisopropyl derivative **3a** in the solid state (Figure S7B), antisymmetric N–C1–N stretching vibration $\nu_{\mathrm{as(N–C1–N)}}$ and symmetric one $\nu_{\mathrm{s(N–C1–N)}}$ were observed at 1618 and 1547 cm^{-1} (calc. 1619 and 1584 cm^{-1}, respectively). The band of 1377 or 1359 cm^{-1} in the experimental data may be assigned to the C1–C2 stretching vibration $\nu_{\mathrm{C1–C2}}$ (calc. 1397 cm^{-1}). On the other hand, $\nu_{\mathrm{as(N–C1–N)}}$ and $\nu_{\mathrm{s(N–C1–N)}}$ in di-*p*-tolyl derivative **4a** were observed at 1594 and 1523 cm^{-1} (calc. 1614 and 1521 cm^{-1}, respectively) in the solid state (Figure S9B). The $\nu_{\mathrm{C1–C2}}$ vibration of **4a** (exp. 1374 cm^{-1}, calc. 1371 cm^{-1}) was also observed at similar position to the case of **3a**. In contrast, the $\nu_{\mathrm{C1–C2}}$ vibration of diene **5** in the solid state was observed at significantly higher wavenumber side (exp. 1515 cm^{-1}, calc. 1529 cm^{-1}) (Figure S11B). In the light of typical absorption positions for C–C single bond (~1000 cm^{-1}) and C=C double bond (~1640 cm^{-1}), the $\nu_{\mathrm{C1–C2}}$ values in **3a** and **4a** indicate a considerable level of the single-bond character. The 'C=C(Tf_2)' bond in **5** also shows electron-deficient but still pronounced double-bond character. It supports the results of the solid state-IR studies that IR spectra of push-pull ethylenes **3a** and **4a** in $CHCl_3$ were essentially similar to those of the solid state (Figures S7C, S9C, and S11C). These data are totally consistent to our conclusions on the basis of XWR-derived bond analysis and NMR analysis.

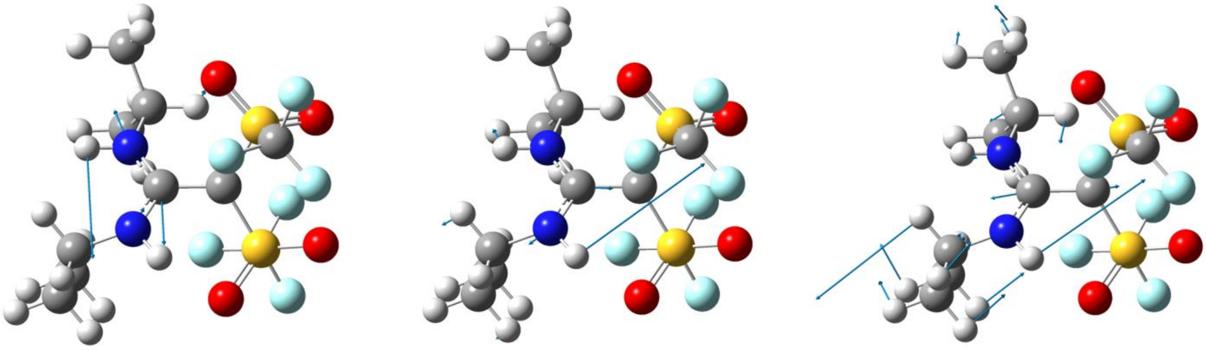


(A) Experimental IR and Raman spectra in solid



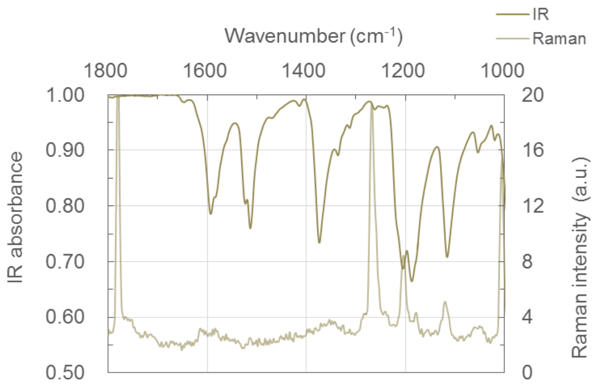
(B) Experimental and theoretical IR spectra in solid (C) Experimental and theoretical IR spectra in $CHCl_3$

Figure S7. IR and Raman spectra of **3a**. Theoretical IR spectra are corrected by scaling factor of 0.975.

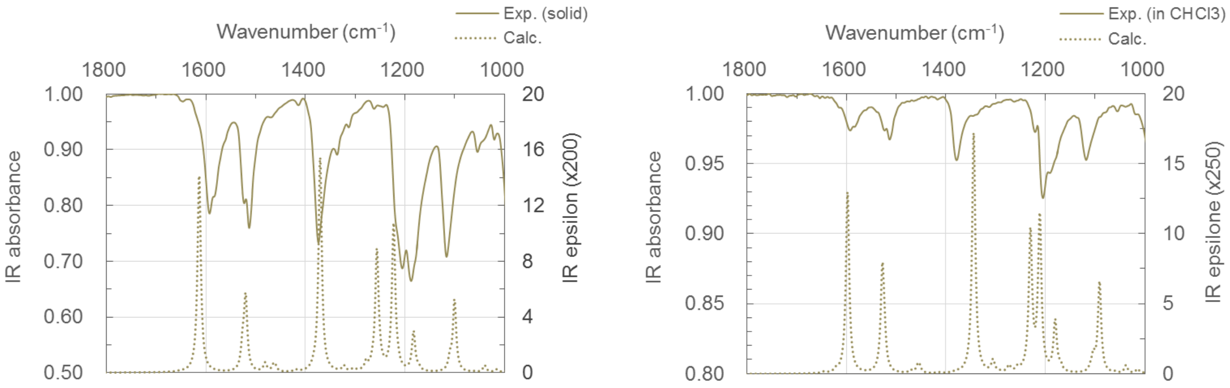


(A) for 1619 cm^{-1} (B) for 1584 cm^{-1} (C) for 1397 cm^{-1}

Figure S8. Frequency vectors obtained by vibration analysis of **3a**

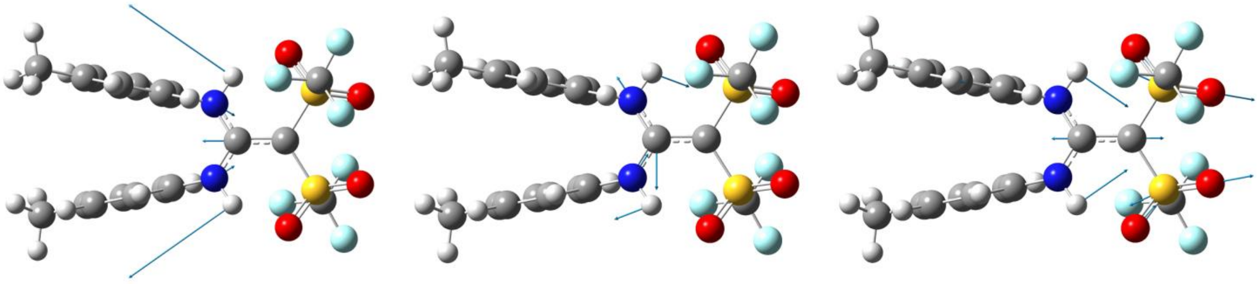


(A) Experimental IR and Raman spectra in solid



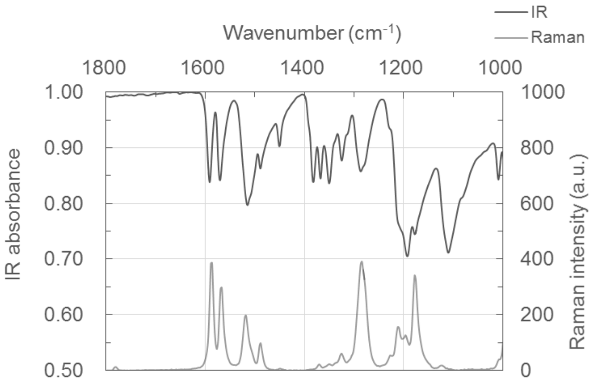
(B) Experimental and theoretical IR spectra in solid (C) Experimental and theoretical IR spectra in $CHCl_3$

Figure S9. IR and Raman spectra of **4a**. Theoretical IR spectra are corrected by scaling factor of 0.975.

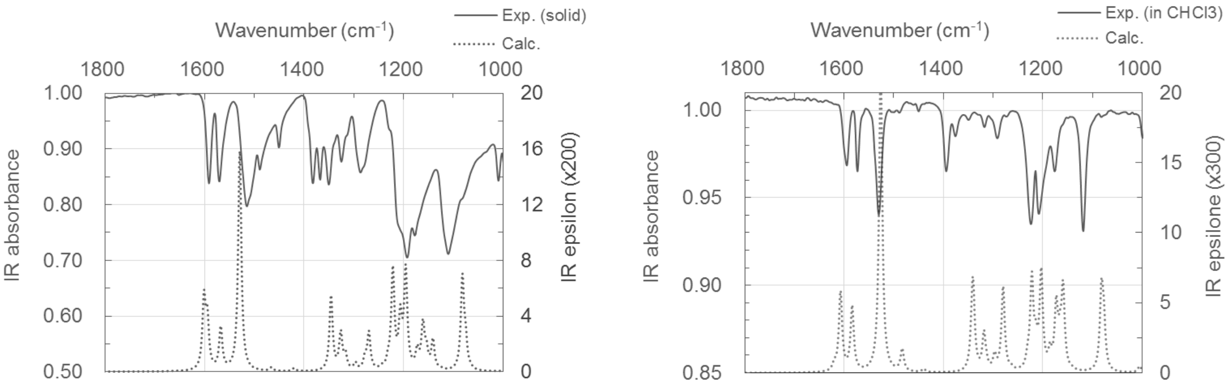


(A) for 1614 cm^{-1} (B) for 1521 cm^{-1} (C) for 1371 cm^{-1}

Figure S10. Frequency vectors obtained by vibration analysis of **4a**

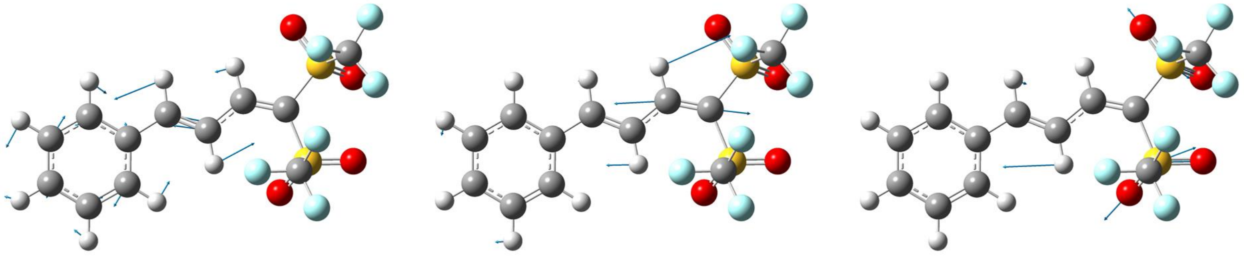


(A) Experimental IR and Raman spectra in solid



(B) Experimental and theoretical IR spectra in solid (C) Experimental and theoretical IR spectra in $CHCl_3$

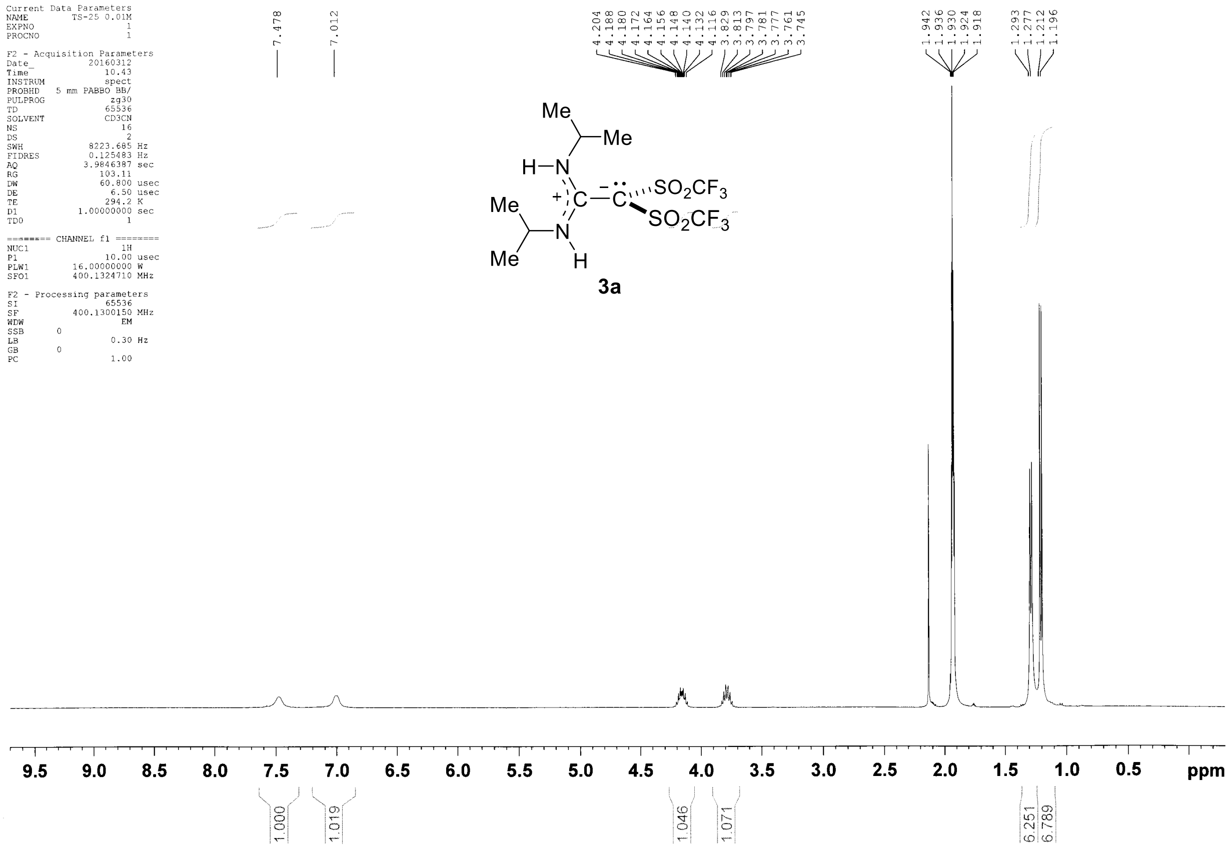
Figure S11. IR and Raman spectra of **5**. Theoretical IR spectra are corrected by scaling factor of 0.950.



(A) for 1601 cm^{-1} (B) for 1529 cm^{-1} (C) for 1345 cm^{-1}

Figure S12. Frequency vectors obtained by vibration analysis of **5**.

6. ^{1}H and ^{13}C NMR spectra of all products



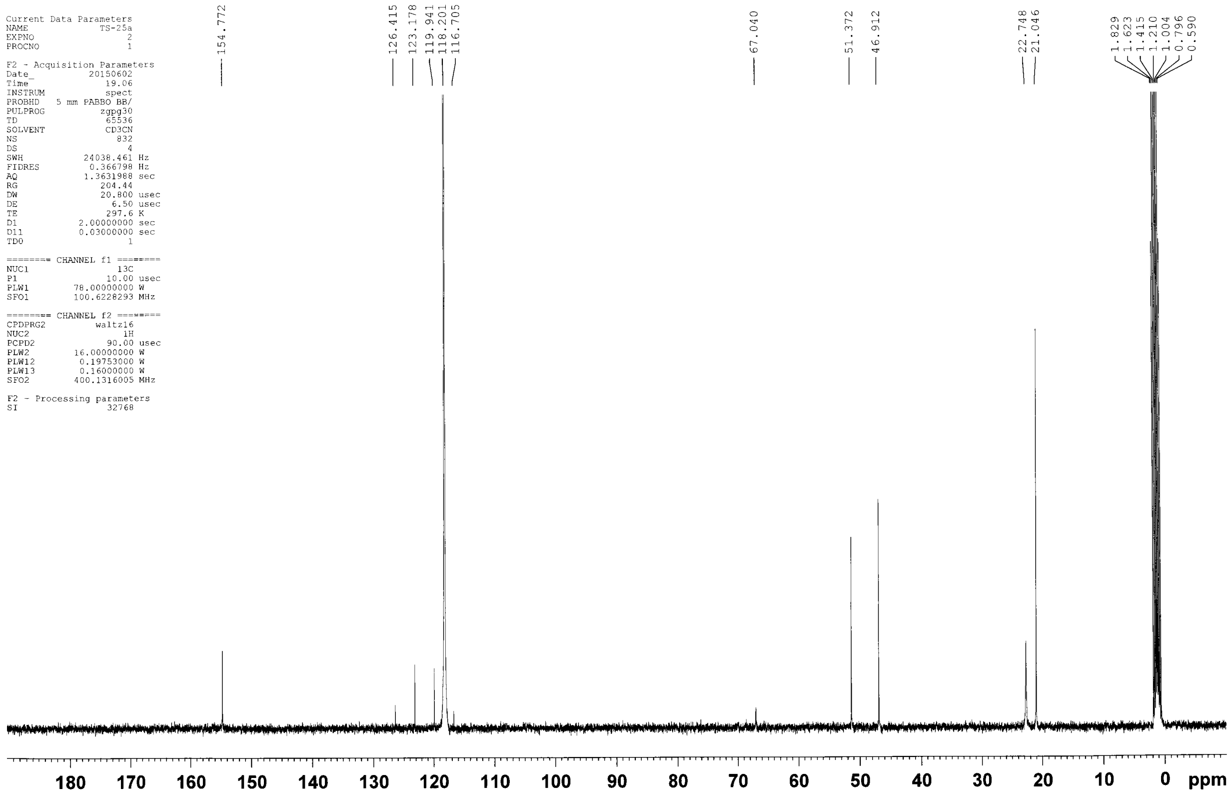
```
Current Data Parameters
NAME         TS-25 0.01M
EXPNO                  1
PROCNO                 1

F2 - Acquisition Parameters
Date_           20160312
Time               10.43
INSTRUM            spect
PROBHD   5 mm PABBO BB/
PULPROG             zg30
TD                 65536
SOLVENT            CD3CN
NS                    16
DS                     2
SWH             8223.685 Hz
FIDRES          0.125483 Hz
AQ             3.9846387 sec
RG                103.11
DW                60.800 usec
DE                  6.50 usec
TE                 294.2 K
D1            1.00000000 sec
TD0                    1

======== CHANNEL f1 ========
NUC1                  1H
P1                 10.00 usec
PLW1         16.00000000 W
SFO1         400.1324710 MHz

F2 - Processing parameters
SI                 65536
SF           400.1300150 MHz
WDW                   EM
SSB      0
LB                  0.30 Hz
GB       0
PC                  1.00
```

3a



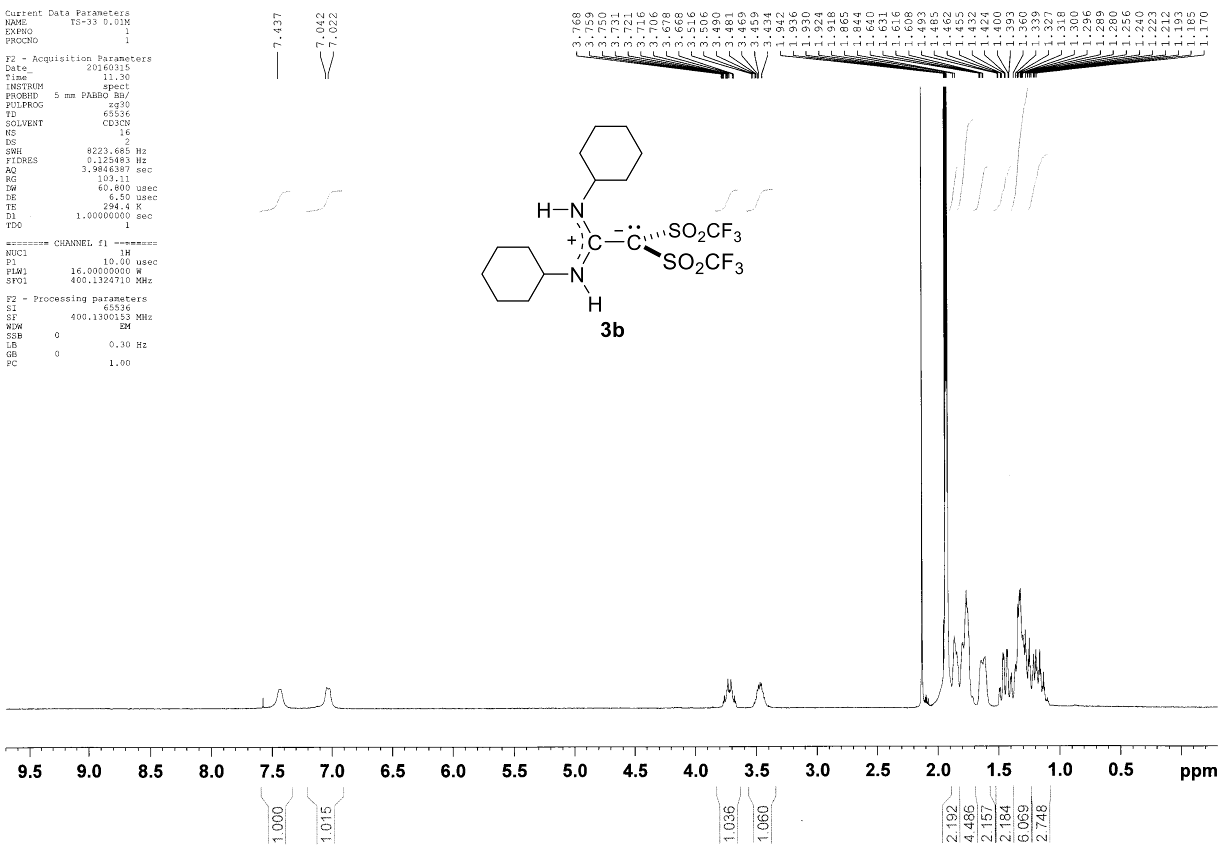
```
Current Data Parameters
NAME               TS-25a
EXPNO                   2
PROCNO                  1

F2 - Acquisition Parameters
Date_            20150602
Time                19.06
INSTRUM             spect
PROBHD   5 mm PABBO BB/
PULPROG           zgpg30
TD                  65536
SOLVENT             CD3CN
NS                    832
DS                      4
SWH             24038.461 Hz
FIDRES           0.366798 Hz
AQ              1.3631988 sec
RG                 204.44
DW                 20.800 usec
DE                   6.50 usec
TE                  297.6 K
D1             2.00000000 sec
D11            0.03000000 sec
TD0                     1

======== CHANNEL f1 ========
NUC1                  13C
P1                  10.00 usec
PLW1          78.00000000 W
SFO1          100.6228293 MHz

======== CHANNEL f2 ========
CPDPRG2           waltz16
NUC2                   1H
PCPD2               90.00 usec
PLW2          16.00000000 W
PLW12          0.19753000 W
PLW13          0.16000000 W
SFO2          400.1316005 MHz

F2 - Processing parameters
SI                  32768
```

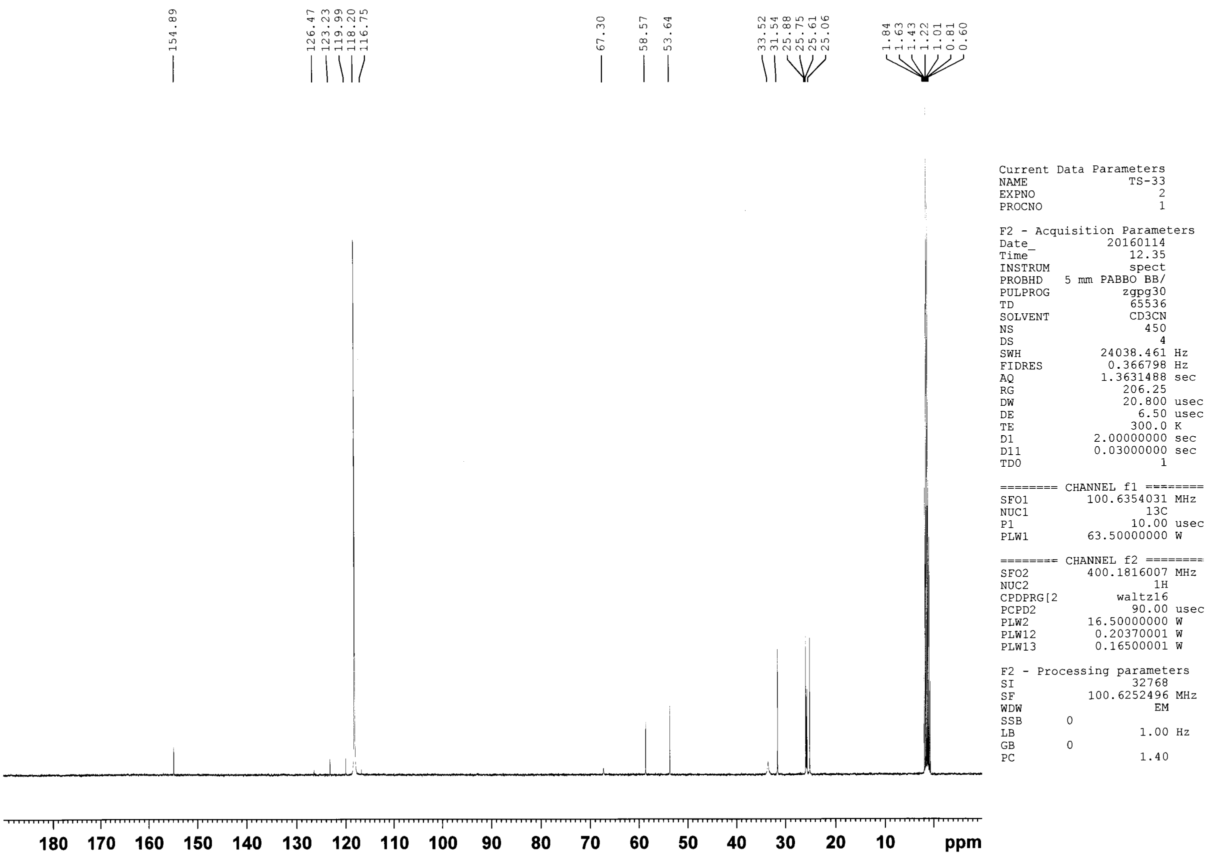


```
Current Data Parameters
NAME          TS-33 0.01M
EXPNO                   1
PROCNO                  1

F2 - Acquisition Parameters
Date_            20160315
Time                11.30
INSTRUM             spect
PROBHD   5 mm PABBO BB/
PULPROG              zg30
TD                  65536
SOLVENT             CD3CN
NS                     16
DS                      2
SWH              8223.685 Hz
FIDRES           0.125483 Hz
AQ              3.9846387 sec
RG                 103.11
DW                 60.800 usec
DE                   6.50 usec
TE                  294.4 K
D1             1.00000000 sec
TD0                     1

======== CHANNEL f1 ========
NUC1                   1H
P1                  10.00 usec
PLW1          16.00000000 W
SFO1          400.1324710 MHz

F2 - Processing parameters
SI                  65536
SF            400.1300153 MHz
WDW                    EM
SSB      0
LB                   0.30 Hz
GB       0
PC                   1.00
```



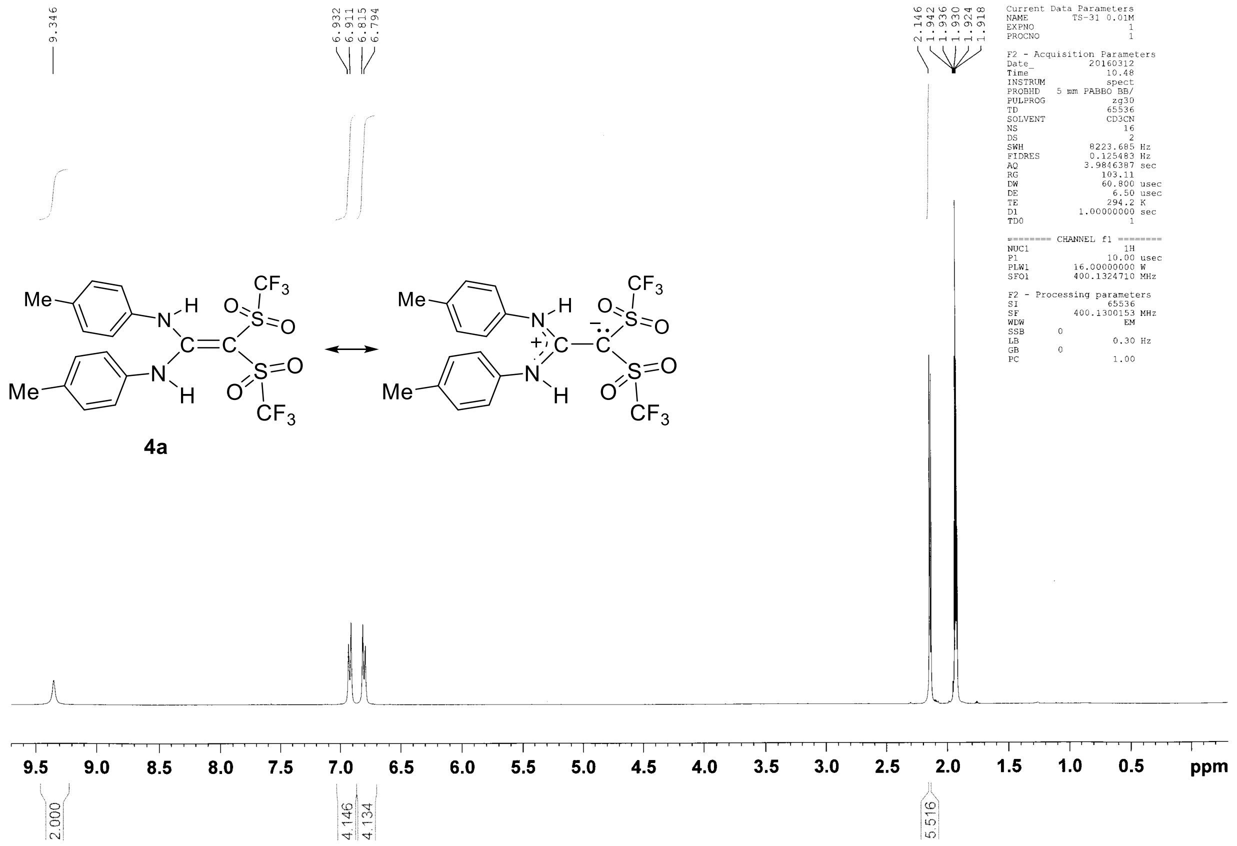
```
Current Data Parameters
NAME                TS-33
EXPNO                   2
PROCNO                  1

F2 - Acquisition Parameters
Date_            20160114
Time                12.35
INSTRUM             spect
PROBHD   5 mm PABBO BB/
PULPROG            zgpg30
TD                  65536
SOLVENT             CD3CN
NS                    450
DS                      4
SWH             24038.461 Hz
FIDRES           0.366798 Hz
AQ              1.3631488 sec
RG                 206.25
DW                 20.800 usec
DE                   6.50 usec
TE                  300.0 K
D1             2.00000000 sec
D11            0.03000000 sec
TD0                     1

======== CHANNEL f1 ========
SFO1          100.6354031 MHz
NUC1                  13C
P1                  10.00 usec
PLW1          63.50000000 W

======== CHANNEL f2 ========
SFO2          400.1816007 MHz
NUC2                   1H
CPDPRG[2          waltz16
PCPD2               90.00 usec
PLW2          16.50000000 W
PLW12          0.20370001 W
PLW13          0.16500001 W

F2 - Processing parameters
SI                  32768
SF            100.6252496 MHz
WDW                    EM
SSB      0
LB                   1.00 Hz
GB       0
PC                   1.40
```



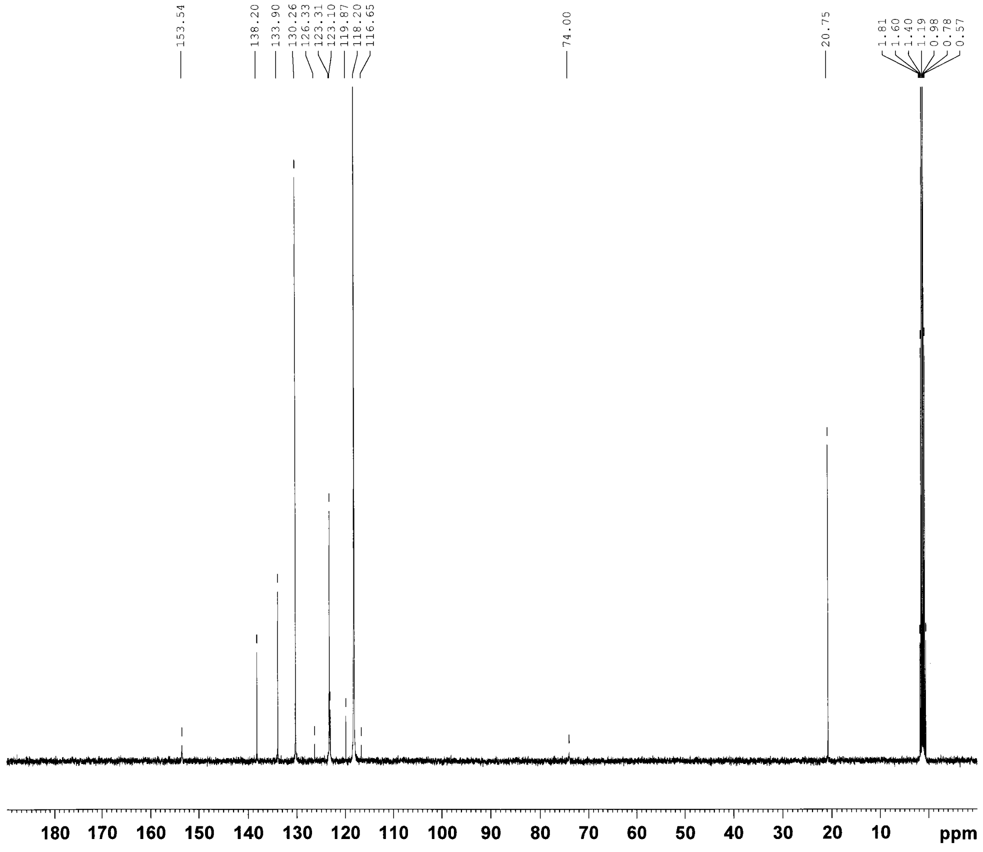
4a

```
Current Data Parameters
NAME          TS-31 0.01M
EXPNO                   1
PROCNO                  1

F2 - Acquisition Parameters
Date_            20160312
Time                10.48
INSTRUM             spect
PROBHD   5 mm PABBO BB/
PULPROG              zg30
TD                  65536
SOLVENT             CD3CN
NS                     16
DS                      2
SWH              8223.685 Hz
FIDRES           0.125483 Hz
AQ              3.9846387 sec
RG                 103.11
DW                 60.800 usec
DE                   6.50 usec
TE                  294.2 K
D1             1.00000000 sec
TD0                     1

======== CHANNEL f1 ========
NUC1                   1H
P1                  10.00 usec
PLW1          16.00000000 W
SFO1          400.1324710 MHz

F2 - Processing parameters
SI                  65536
SF            400.1300153 MHz
WDW                    EM
SSB      0
LB                   0.30 Hz
GB       0
PC                   1.00
```



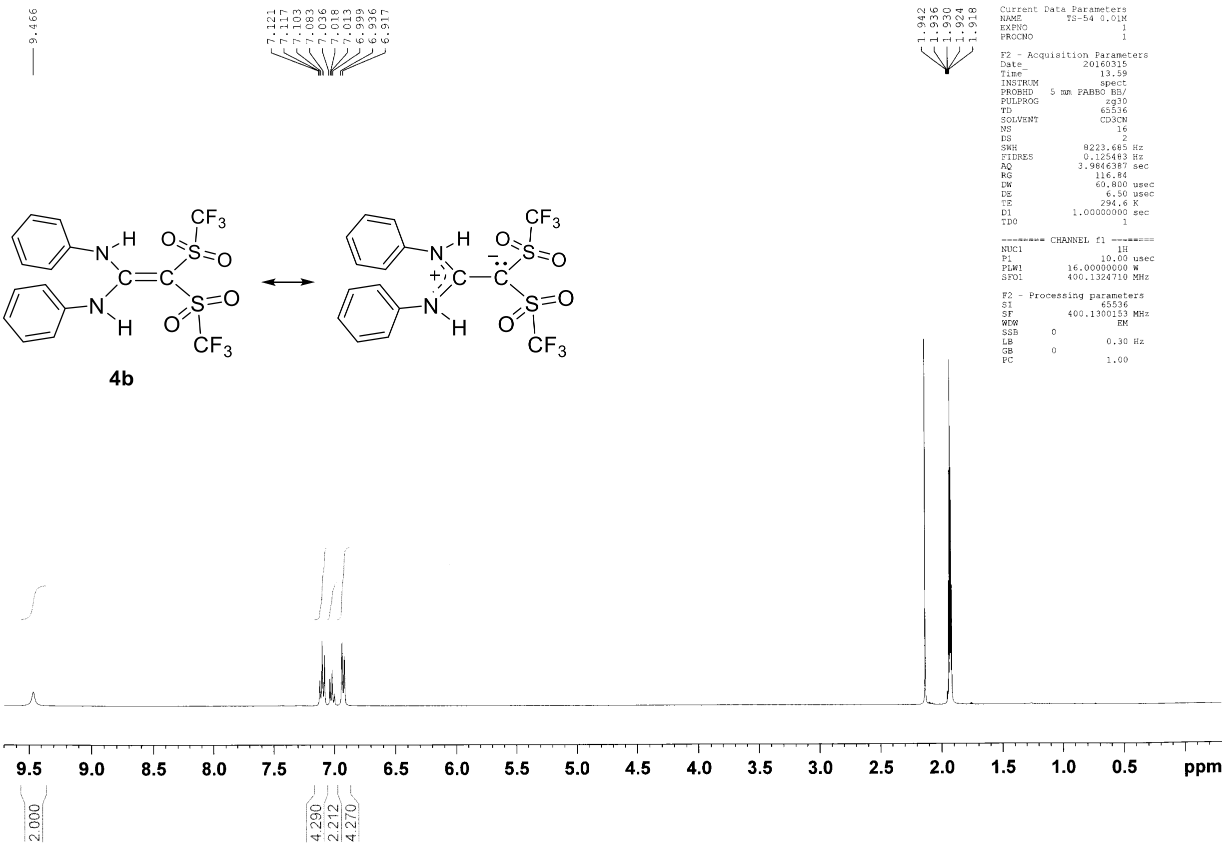
```
Current Data Parameters
NAME             TS-31 re
EXPNO                   2
PROCNO                  1

F2 - Acquisition Parameters
Date_            20160113
Time                20.26
INSTRUM             spect
PROBHD   5 mm PABBO BB/
PULPROG           zgpg30
TD                  65536
SOLVENT             CD3CN
NS                    173
DS                      4
SWH             24038.461 Hz
FIDRES           0.366798 Hz
AQ              1.3631988 sec
RG                 204.44
DW                 20.800 usec
DE                   6.50 usec
TE                  294.9 K
D1             2.00000000 sec
D11            0.03000000 sec
TD0                     1

======== CHANNEL f1 ========
NUC1                  13C
P1                  10.00 usec
PLW1          78.00000000 W
SFO1          100.6228293 MHz

======== CHANNEL f2 ========
CPDPRG2           waltz16
NUC2                   1H
PCPD2               90.00 usec
PLW2          16.00000000 W
PLW12          0.19753000 W
PLW13          0.16000000 W
SFO2          400.1316005 MHz

F2 - Processing parameters
SI                  32768
SF            100.6126835 MHz
WDW                    EM
SSB      0
LB                   1.00 Hz
GB       0
PC                   1.40
```



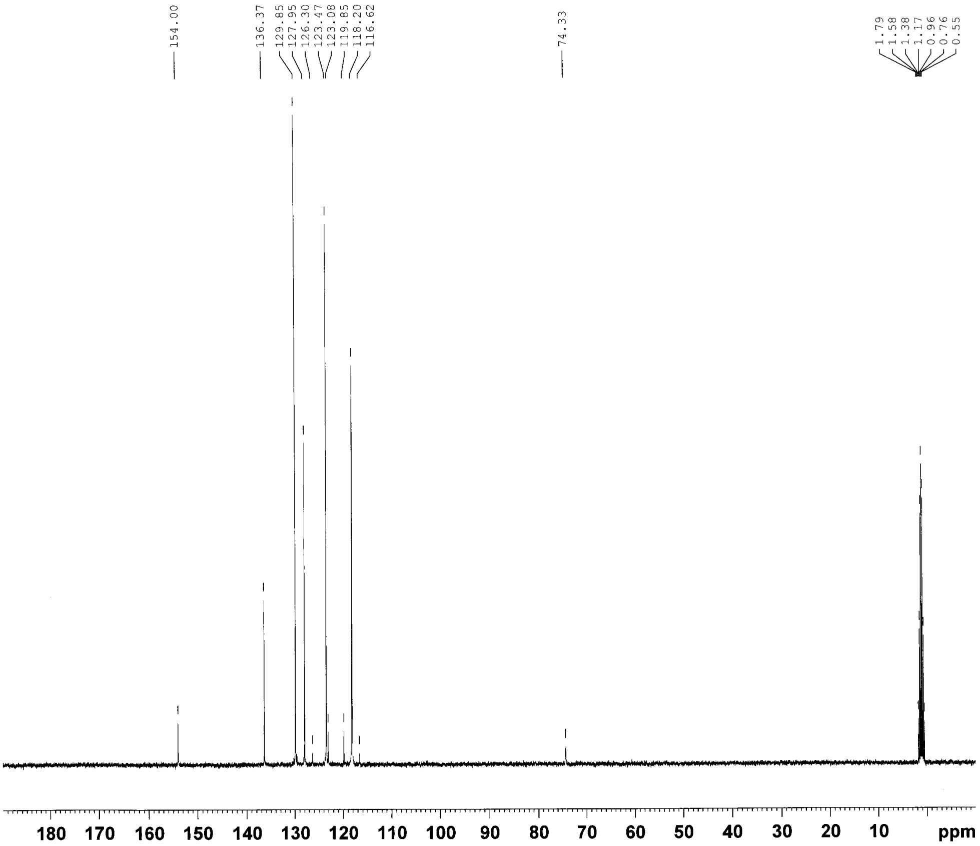
4b

```
Current Data Parameters
NAME          TS-54 0.01M
EXPNO                   1
PROCNO                  1

F2 - Acquisition Parameters
Date_            20160315
Time                13.59
INSTRUM             spect
PROBHD   5 mm PABBO BB/
PULPROG              zg30
TD                  65536
SOLVENT             CD3CN
NS                     16
DS                      2
SWH              8223.685 Hz
FIDRES           0.125483 Hz
AQ              3.9846387 sec
RG                 116.84
DW                 60.800 usec
DE                   6.50 usec
TE                  294.6 K
D1             1.00000000 sec
TD0                     1

======== CHANNEL f1 ========
NUC1                   1H
P1                  10.00 usec
PLW1          16.00000000 W
SFO1          400.1324710 MHz

F2 - Processing parameters
SI                  65536
SF            400.1300153 MHz
WDW                    EM
SSB      0
LB                   0.30 Hz
GB       0
PC                   1.00
```



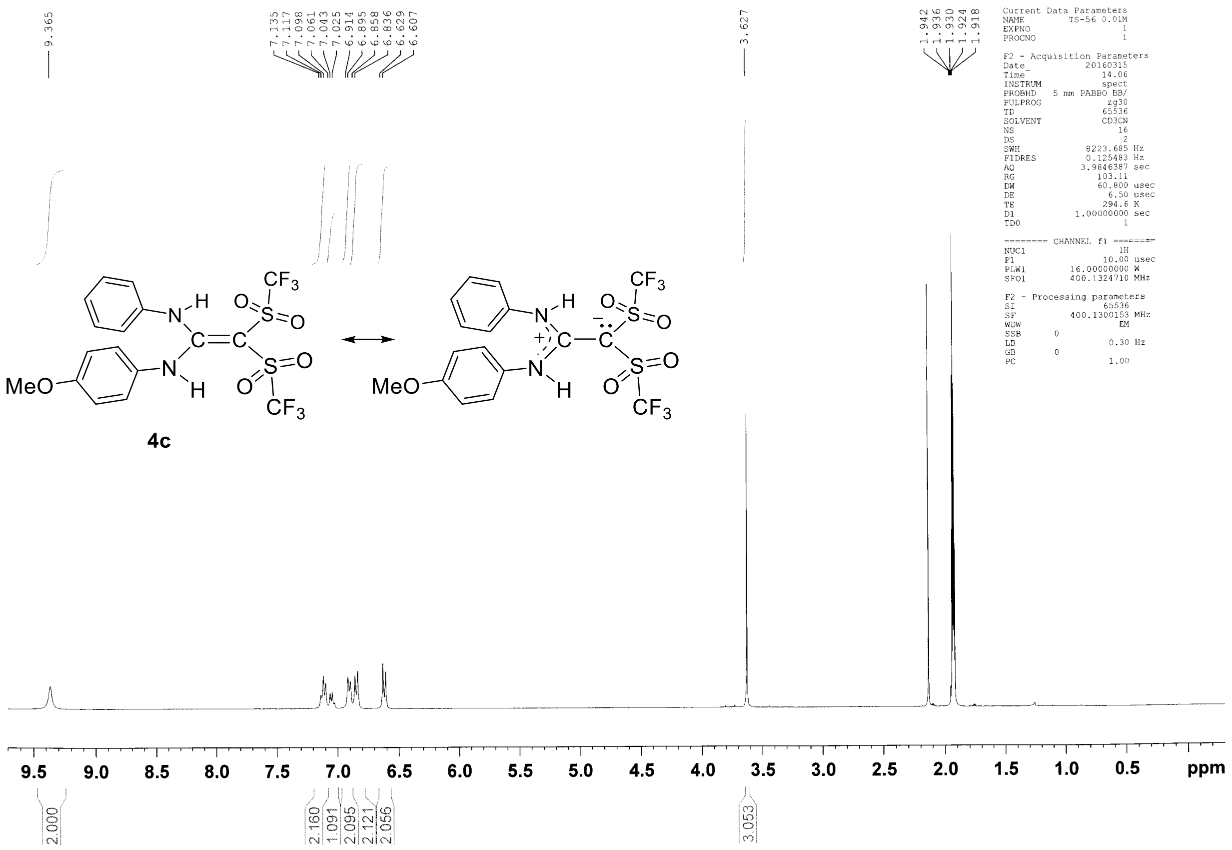
```
Current Data Parameters
NAME          TS-54 CD3CN
EXPNO                   3
PROCNO                  1

F2 - Acquisition Parameters
Date_            20160204
Time                19.33
INSTRUM             spect
PROBHD   5 mm PABBO BB/
PULPROG            zgpg30
TD                  65536
SOLVENT             CD3CN
NS                     76
DS                      4
SWH             24038.461 Hz
FIDRES           0.366798 Hz
AQ              1.3631988 sec
RG                 204.44
DW                 20.800 usec
DE                   6.50 usec
TE                  294.8 K
D1             2.00000000 sec
D11            0.03000000 sec
TD0                     1

======== CHANNEL f1 ========
NUC1                  13C
P1                  10.00 usec
PLW1          78.00000000 W
SFO1          100.6228293 MHz

======== CHANNEL f2 ========
CPDPRG2           waltz16
NUC2                   1H
PCPD2               90.00 usec
PLW2          16.00000000 W
PLW12          0.19753000 W
PLW13          0.16000000 W
SFO2          400.1316005 MHz

F2 - Processing parameters
SI                  32768
SF            100.6126877 MHz
WDW                    EM
SSB      0
LB                   1.00 Hz
GB       0
PC                   1.40
```



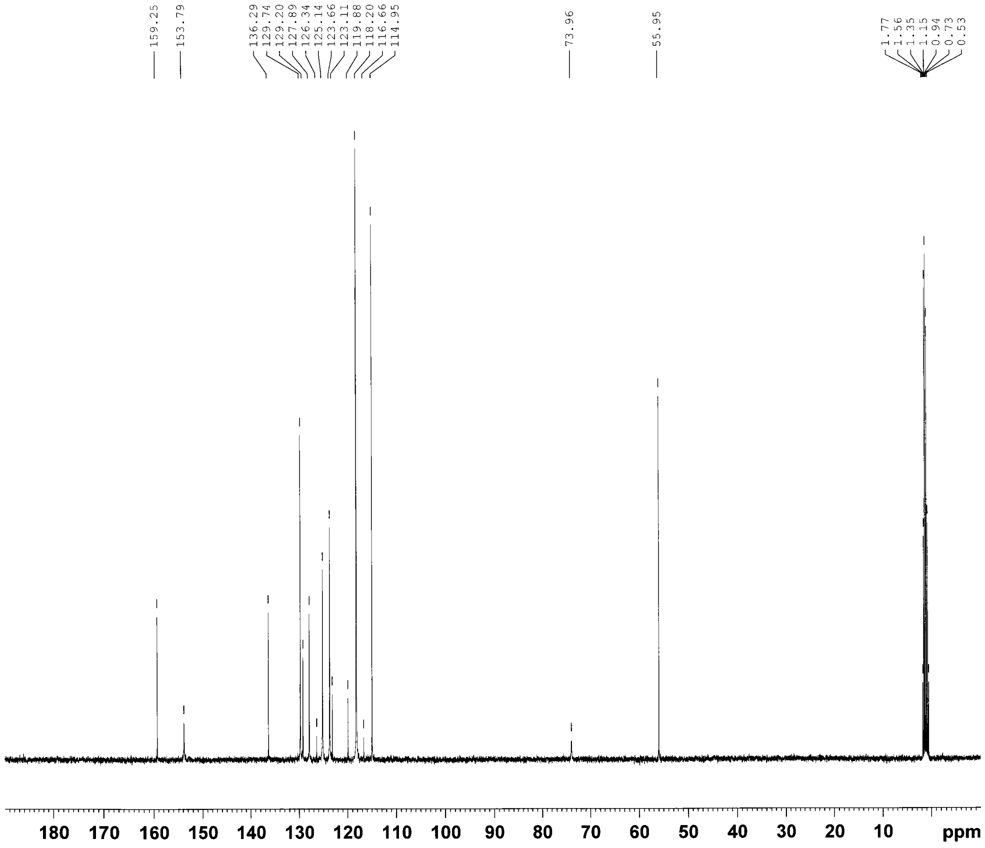
4c

```
Current Data Parameters
NAME          TS-56 0.01M
EXPNO                   1
PROCNO                  1

F2 - Acquisition Parameters
Date_            20160315
Time                14.06
INSTRUM             spect
PROBHD   5 mm PABBO BB/
PULPROG              zg30
TD                  65536
SOLVENT             CD3CN
NS                     16
DS                      2
SWH              8223.685 Hz
FIDRES           0.125483 Hz
AQ              3.9846387 sec
RG                 103.11
DW                 60.800 usec
DE                   6.50 usec
TE                  294.6 K
D1             1.00000000 sec
TD0                     1

======== CHANNEL f1 ========
NUC1                   1H
P1                  10.00 usec
PLW1          16.00000000 W
SFO1          400.1324710 MHz

F2 - Processing parameters
SI                  65536
SF            400.1300153 MHz
WDW                    EM
SSB      0
LB                   0.30 Hz
GB       0
PC                   1.00
```



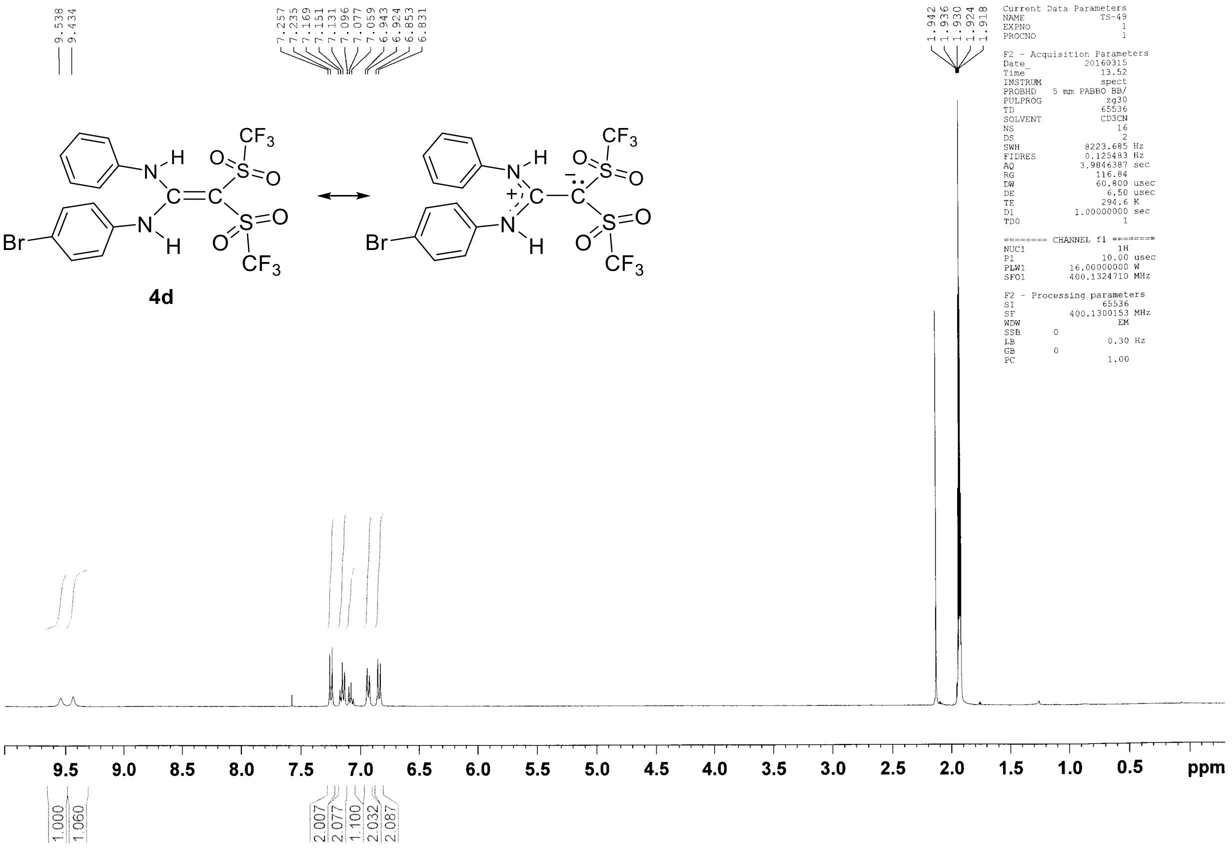
```
Current Data Parameters
NAME          TS-56 CD3CN
EXPNO                   3
PROCNO                  1

F2 - Acquisition Parameters
Date_            20160204
Time                16.01
INSTRUM             spect
PROBHD   5 mm PABBO BB/
PULPROG            zgpg30
TD                  65536
SOLVENT             CD3CN
NS                     79
DS                      4
SWH             24038.461 Hz
FIDRES           0.366798 Hz
AQ              1.3631988 sec
RG                 204.44
DW                 20.800 usec
DE                   6.50 usec
TE                  294.9 K
D1             2.00000000 sec
D11            0.03000000 sec
TD0                     1

======== CHANNEL f1 ========
NUC1                  13C
P1                  10.00 usec
PLW1          78.00000000 W
SFO1          100.6228293 MHz

======== CHANNEL f2 ========
CPDPRG2           waltz16
NUC2                   1H
PCPD2               90.00 usec
PLW2          16.00000000 W
PLW12          0.19753000 W
PLW13          0.16000000 W
SFO2          400.1316005 MHz

F2 - Processing parameters
SI                  32768
SF            100.6126887 MHz
WDW                    EM
SSB      0
LB                   1.00 Hz
GB       0
PC                   1.40
```



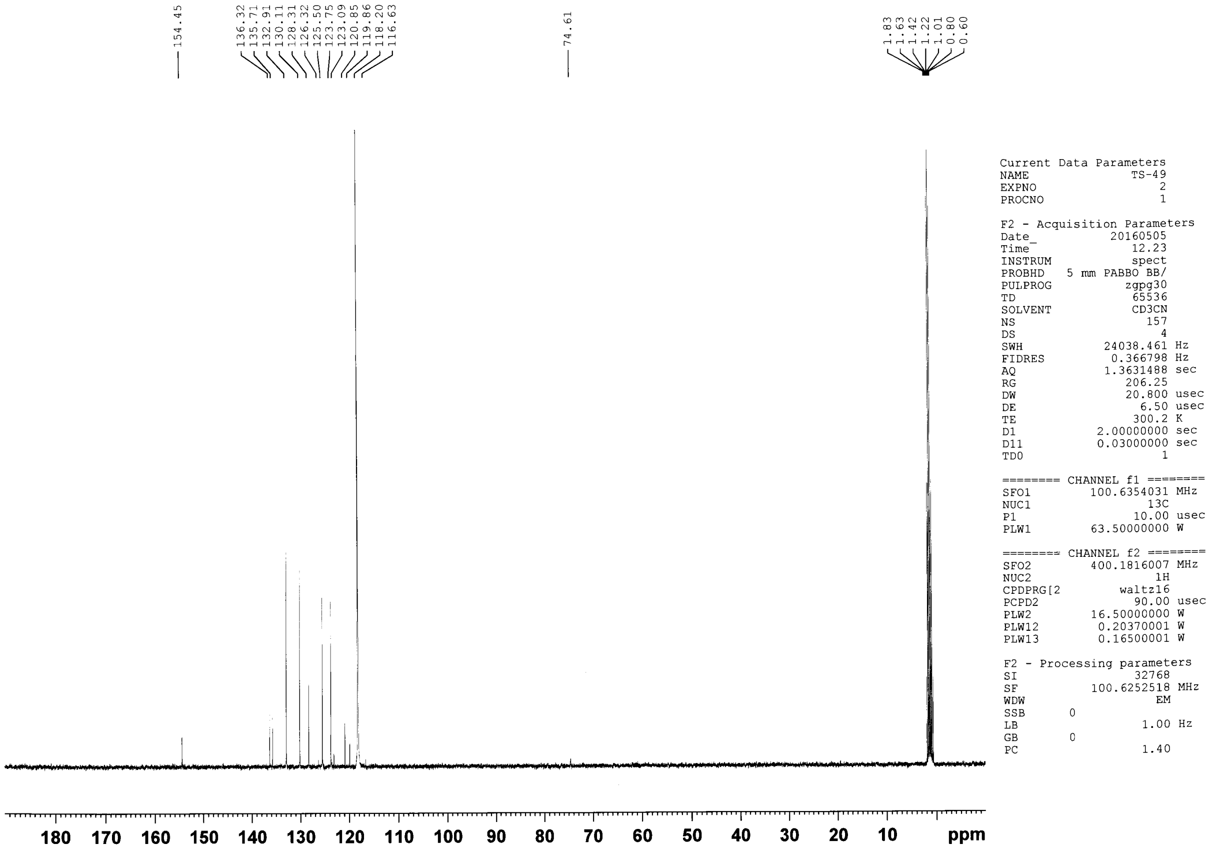
4d

```
Current Data Parameters
NAME              TS-49
EXPNO                 1
PROCNO                1

F2 - Acquisition Parameters
Date_          20160315
Time              13.52
INSTRUM           spect
PROBHD   5 mm PABBO BB/
PULPROG            zg30
TD                65536
SOLVENT           CD3CN
NS                   16
DS                    2
SWH            8223.685 Hz
FIDRES         0.125483 Hz
AQ            3.9846387 sec
RG               116.84
DW               60.800 usec
DE                 6.50 usec
TE                294.6 K
D1           1.00000000 sec
TD0                   1

======== CHANNEL f1 ========
NUC1                 1H
P1                10.00 usec
PLW1        16.00000000 W
SFO1        400.1324710 MHz

F2 - Processing parameters
SI                65536
SF          400.1300153 MHz
WDW                  EM
SSB      0
LB                 0.30 Hz
GB       0
PC                 1.00
```



```
Current Data Parameters
NAME              TS-49
EXPNO                 2
PROCNO                1

F2 - Acquisition Parameters
Date_          20160505
Time              12.23
INSTRUM           spect
PROBHD   5 mm PABBO BB/
PULPROG          zgpg30
TD                65536
SOLVENT           CD3CN
NS                  157
DS                    4
SWH           24038.461 Hz
FIDRES         0.366798 Hz
AQ            1.3631488 sec
RG               206.25
DW               20.800 usec
DE                 6.50 usec
TE                300.2 K
D1           2.00000000 sec
D11          0.03000000 sec
TD0                   1

======== CHANNEL f1 ========
SFO1        100.6354031 MHz
NUC1                13C
P1                10.00 usec
PLW1        63.50000000 W

======== CHANNEL f2 ========
SFO2        400.1816007 MHz
NUC2                 1H
CPDPRG[2        waltz16
PCPD2             90.00 usec
PLW2        16.50000000 W
PLW12        0.20370001 W
PLW13        0.16500001 W

F2 - Processing parameters
SI                32768
SF          100.6252518 MHz
WDW                  EM
SSB      0
LB                 1.00 Hz
GB       0
PC                 1.40
```

7. References

1. (a) F. J. Waller, A. G. M. Barrett, D. C. Braddock, D. Ramprasad, R. M. McKinnell, A. J. P. White, D. J. Williams, R. J. Ducray, *J. Org. Chem.* **1999**, *64*, 2910. (b) H. Yanai, A. Takahashi, T. Taguchi, *Tetrahedron* **2007**, *63*, 12149.
2. M. Hanack, J. Hackenberg, O. Menke, L. R. Subramanian, R. Schlichenmaier *Synthesis* **1994**, 249.
3. A. R. Ali, H. Ghosh, B. K. Patel, *Tetrahedron Lett.* **2010**, *51*, 1019.
4. D. Jayatilaka, D. J. Grimwood, *International Conference on Computational Science*, Springer Berlin Heidelberg, 2003.
5. For *Gaussian 09*, Revision D.01, M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, T. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J. M. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels, O. Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski, D. J. Fox, Gaussian, Inc., Wallingford CT, 2013.
6. (a) M. C. Durrant, *Chem. Sci.* **2015**, *6*, 6614-6623. (b) M. Fugel, L. A. Malaspina, R. Pal, S. P. Thomas, M. W. Shi, M. A. Spackman, K. Sugimoto, S. Grabowsky, *Chem. Eur. J.* 2019, Accepted Article (DOI: 10.1002/chem.201806247).
7. Y. Zhao, D. G. Truhlar, *Theor. Chem. Acc.* **2008**, *120*: 215–241.
8. For *Firefly 8.0*, A. A. Granovsky, http://classic.chem.msu.su/gran/firefly/index.html
9. For *NBO 6.0*, E. D. Glendening, J, K. Badenhoop, A. E. Reed, J. E. Carpenter, J. A. Bohmann, C. M. Morales, C. R. Landis, F. Weinhold, Theoretical Chemistry Institute, University of Wisconsin, Madison, 2013.
10. For *ADF 2018*, SCM, Theoretical Chemistry, Vrije Universiteit, Amsterdam, The Netherlands, http://www.scm.com.
11. G. te Velde, F. M. Bickelhaupt, E. J. Baerends, C. F. Guerra, S. J. A. van Gisbergen, J. G. Snijders, T. Ziegler, *J. Comput. Chem.* **2001**, *22*, 931.

Appendix D

Applying Quantum Crystallography for Rational Drug Design

D.1 Supplementary Material for *Publication:* Similarities and Differences between Crystal and Enzyme Environmental Effects on the Electron Density of Drug Molecules

SUPPORTING INFORMATION

Table of Contents

1. Diffraction experiment, radiation damage
2. Multipole modeling
3. Comparison between theoretical and experimental multipole model and XWR
4. Differences in bond densities

1. Diffraction experiment, radiation damage

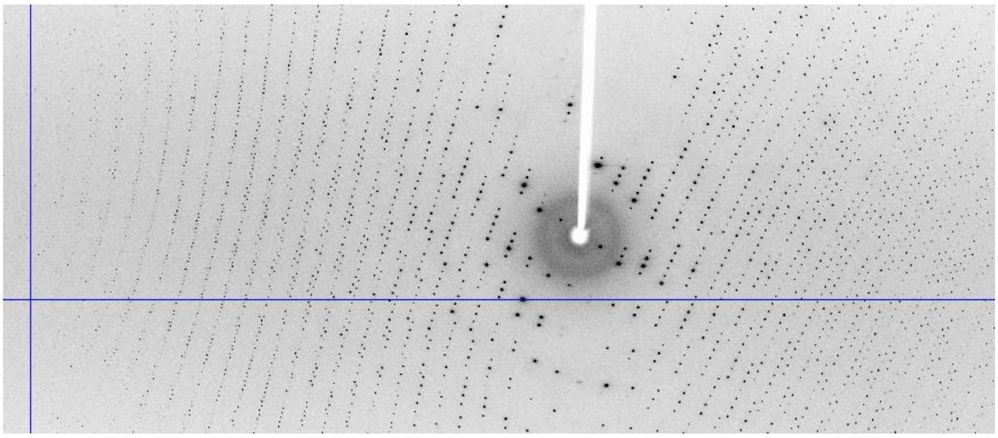


Figure S1: First detection image of **1**K·H_2O measured at BL02B1 of SPring-8 with a curved imaging plate at 25 K. The crosshair of the blue lines indicates a resolution of d=0.220 Å.

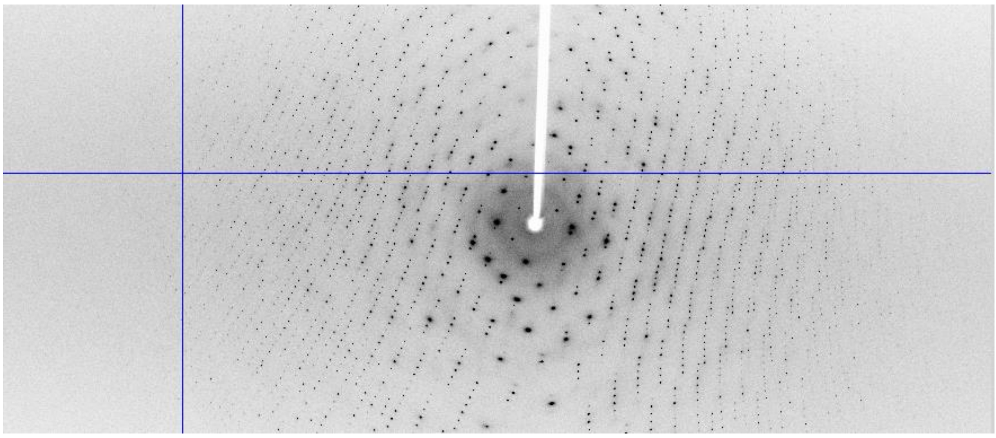


Figure S2: Last detection image of **1**K·H_2O. The crosshair of the blue lines indicates a resolution of d=0.436 Å.

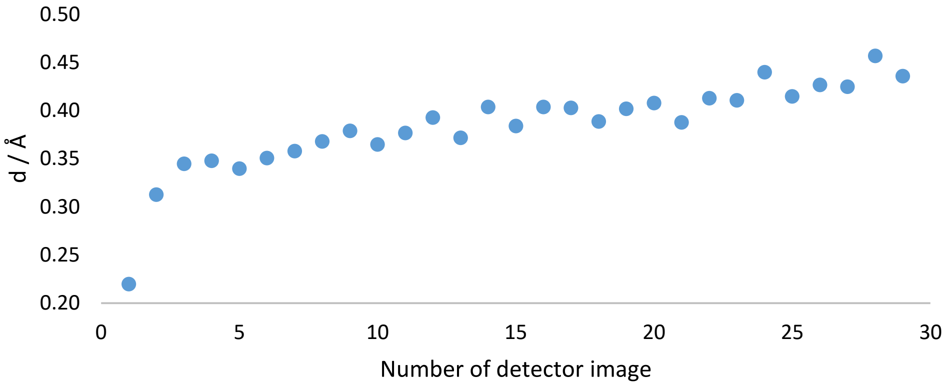


Figure S3: Maximum resolution for the most highly resolved reflection that exceeds the intensity of 190 counts on the imaging plate as a visualization of the radiation damage process.

Table S1: Intensity statistics after correction of radiation damage. The final data set was cut at a maximum resolution of d = 0.45 Å.

Resolution shell (Å)	**No. of measured data**	**Completeness** (%)	**Redundancy**	**R_{merge}**
Inf-1.44	275	100	7,67	0.0207
1.44-0.95	634	100	7,99	0.0267
0.95-0.75	937	100	7,95	0.0309
0.75-0.66	863	100	7,69	0.0381
0.66-0.60	881	100	7,7	0.0440
0.60-0.55	1060	100	7,55	0.0512
0.55-0.52	863	100	7,31	0.0613
0.52-0.49	1070	100	7,13	0.0713
0.49-0.47	868	100	7,12	0.0869
0.47-0.45	992	100	6,63	0.0971
0.45-0.44	624	100	6,54	0.1106
0.44-0.42	1362	100	6,26	0.1308
0.42-0.41	768	100	6,07	0.1529
0.41-0.40	849	100	5,72	0.1679
0.40-0.39	938	99,9	5,72	0.2013
0.39-0.38	1068	99,9	5,62	0.2239
0.38-0.37	1127	99,7	5,29	0.2539
0.37-0.36	1330	99,1	5,25	0.2912
0.36-0.35	1436	99,2	4,95	0.3472

2. Multipole modeling

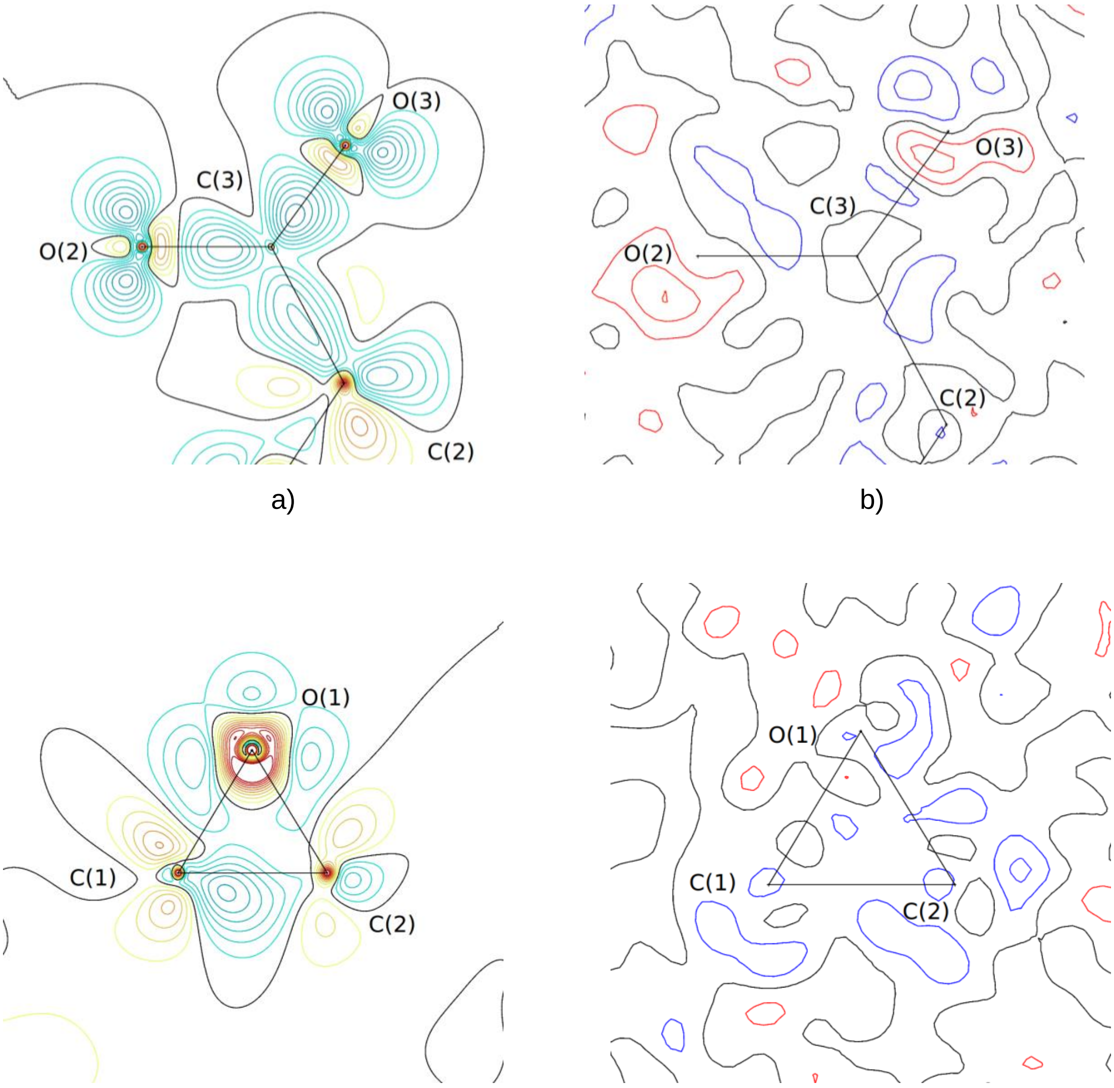
Multipole modeling was carried out in the software XD2006 (A Computer Program Package for Multipole Refinement, Topological Analysis of Charge Densities and Evaluation of Intermolecular Energies from Experimental and Theoretical Structure Factors. A. Volkov, P. Macchi, L. J. Farrugia, C. Gatti, P. Mallinson, T. Richter, T. Koritsanszky, 2006). Local coordinate systems were applied as follows (for the atom numbering see Figure 1 in the main manuscript): C1 and C2 without symmetry constraints (*1*); C3, C4, O2, O3, O4, O5, N1 with mirror symmetry (*m*); O1 with *mm2* symmetry; all H atoms with 6-fold symmetry; K1 and K2 with 2-fold symmetry and further constraints due to their special positions (Table S2). O2 was constrained to O3, H11 to H12. Group charges were defined and used as additional constraints, disallowing charge transfer: asymmetric unit (neutral), anion **1** (-1), water (neutral), each K cation (+0.5). κ-parameters were refined individually for all non-H atoms except for O2 that was constrained to O3 and except for both K-atoms, for which κ was kept at 1.0. κ'-parameters were kept at 1.0 for all non-H atoms. For H-atoms, κ/κ' were chosen as 1.3 and 2.0.

Table S2: Chemical constraints for K atoms due to their special positions

	K1	**K2**
Coordinates kept constant	x, z	z
Further positional constraints	none	x = y
Constraints for ADPs	0,5 U11 = U12 0,5 U13 = U23	U11 = U22 U13 = - U23

During the refinement procedure, the scale factor was refined first, followed by monopoles, dipoles and quadrupoles for all atoms except the spherical K atoms. Subsequently, C-H and N-H bond distances in **1** were elongated to values from the QM/MM geometry optimization in the crystal environment (model **C**) in this study (see main text for details). O-H bond distances in water were elongated to values from neutron diffraction (L. A. Malaspina, A. J. Edwards, M. Woinska, D. Jayatilaka, M. J. Turner, J. R. Price, R. Herbst-Irmer, K. Sugimoto, E. Nishibori, S. Grabowsky, *Cryst. Growth Des.* **2017**, *17*, 3812-3825). The H-atom positions were fixed in subsequent refinements, only their isotropic displacement parameters were further refined. For all C-, O- and N-atoms, octupoles and hexadecapoles as well as anisotropic displacement parameters were refined in the final refinement cycles. Final figures of merit of the multipole refinement are: $R(F)= 0.0246$, $R(F^2)=0.0393$, GooF=3.229.

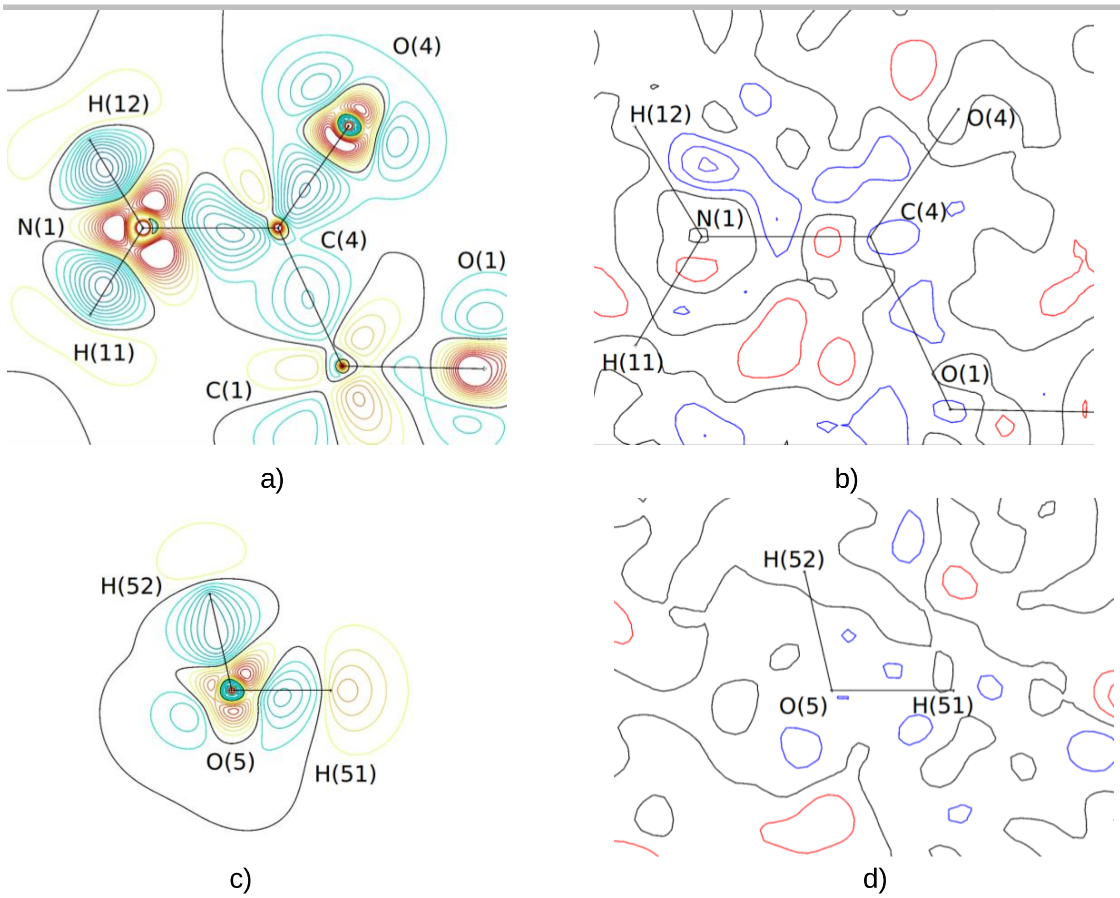
3. Comparison between theoretical and experimental multipole model and XWR



a) b)

c) d)

Figure S4: a),c) Static deformation density maps; b),d) residual electron density maps. a),b) Carboxylate group; c),d) epoxide ring. Based on the multipole refinement against the experimental structure factors as described in Section 2. Blue = positive, red = negative. Black = zero contour line. Contour interval = 0.1 eÅ^{-3}.



a)

b)

c)

d)

Figure S5: a),c) Static deformation density maps; b),d) residual electron density maps. a),b) Amide group; c),d) water molecule. Based on the multipole refinement against the experimental structure factors as described in Section 2. Blue = positive, red = negative. Black = zero contour line. Contour interval = 0.1 eÅ^{-3}.

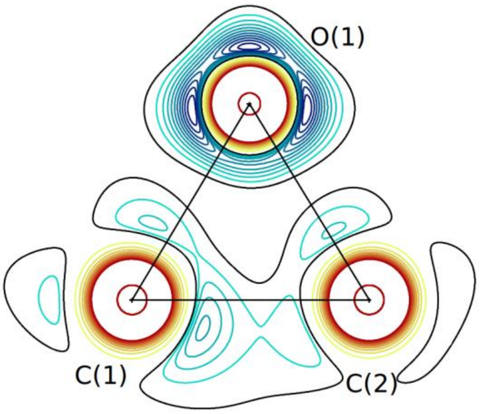


Figure S6: Laplacian map of the epoxide ring. Based on the multipole refinement against the experimental structure factors as described in Section 2. Blue = positive, red = negative. Black = zero contour line. Contour interval = 10 eÅ^{-5} in the positive regions, 100 eÅ^{-5} in the negative regions.

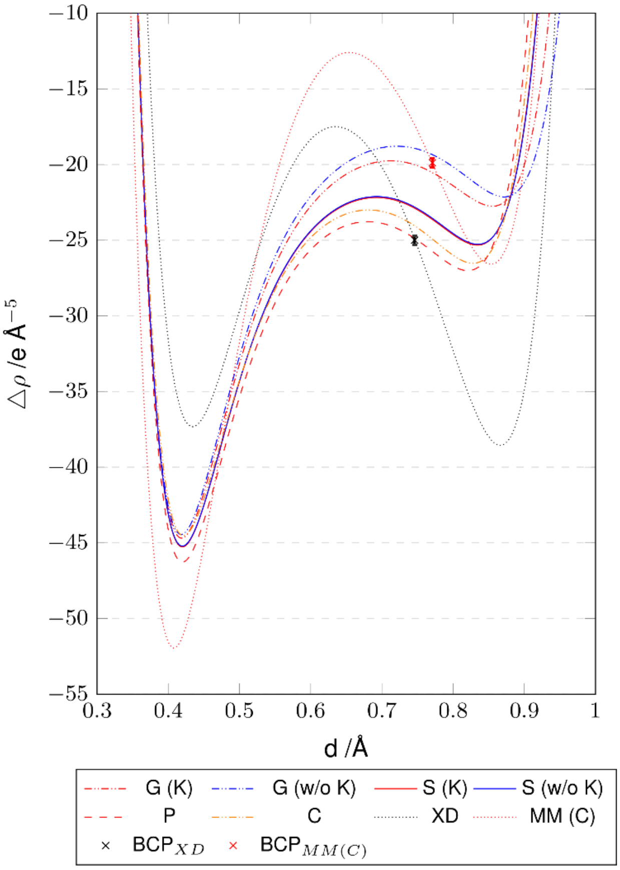


Figure S7. Laplacian of ED along the N1-C4 bond (plotted with the N atom at position 0 Å, left) for the vacuum model (**G**) with or without K$^+$ counter-cation, the solvent model (**S**) with or without K$^+$ counter-cation, the crystal model (**C**), the enzyme model (**P**), an experimental (**XD**) and a theoretical (**MM (C)**) multipole model. The theoretical multipole model is based on synthetic structure factors calculated in a pseudo-periodic environment with the software *Denprop*. The theoretical multipole model deviates as strongly from the theoretical models as the experimental multipole model. BCP = bond critical point.

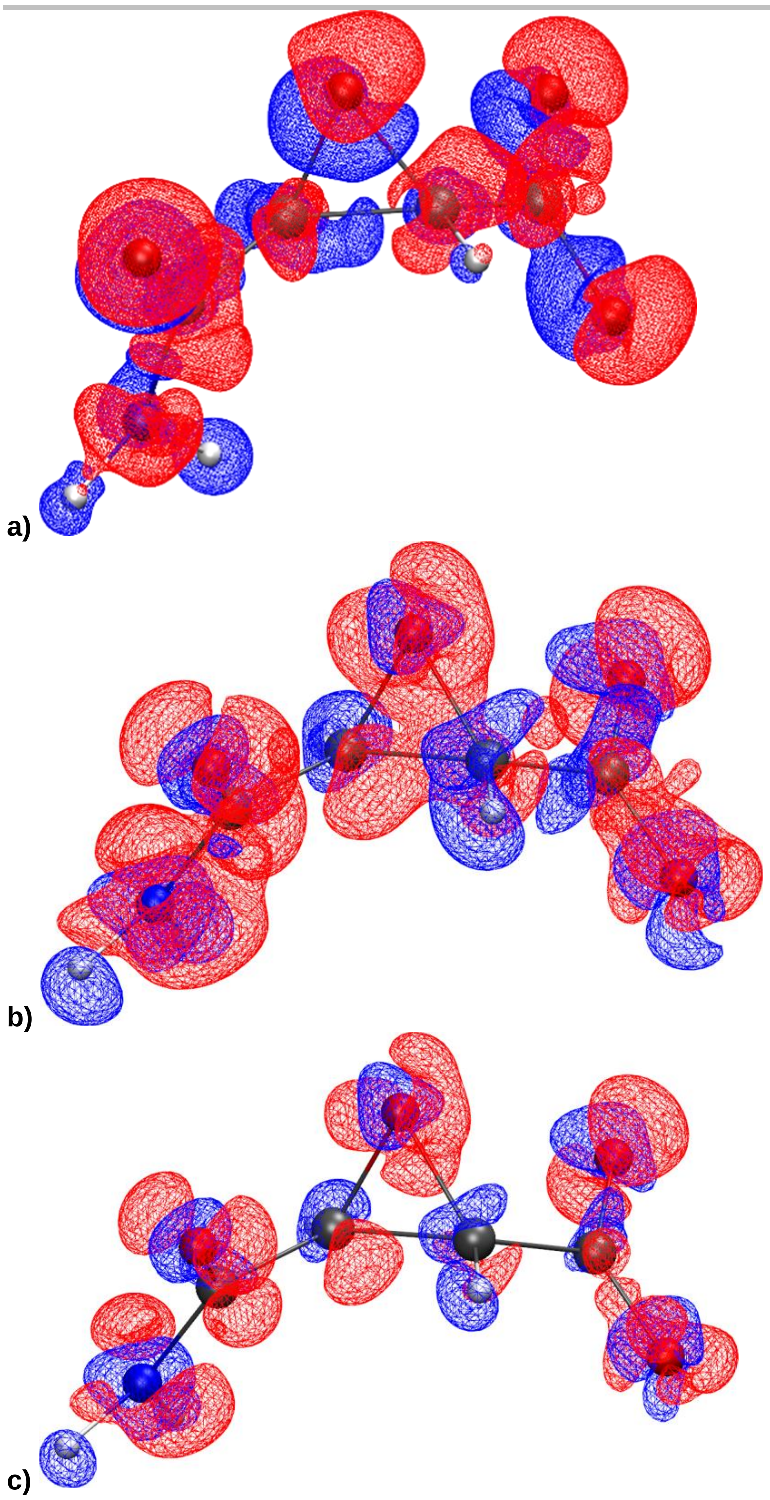


Figure S8. Isosurfaces of the interaction density for a) the experimental multipole model (isovalue ±0.034 *e* Å^{-3}), b) the XCW fitting (isovalue ±0.067 *e* Å^{-3}), c) the XCW fitting (isovalue ±0.134 *e* Å^{-3}). Red regions correspond to higher ED compared to the *in-vacuo* state, blue to lower ED.

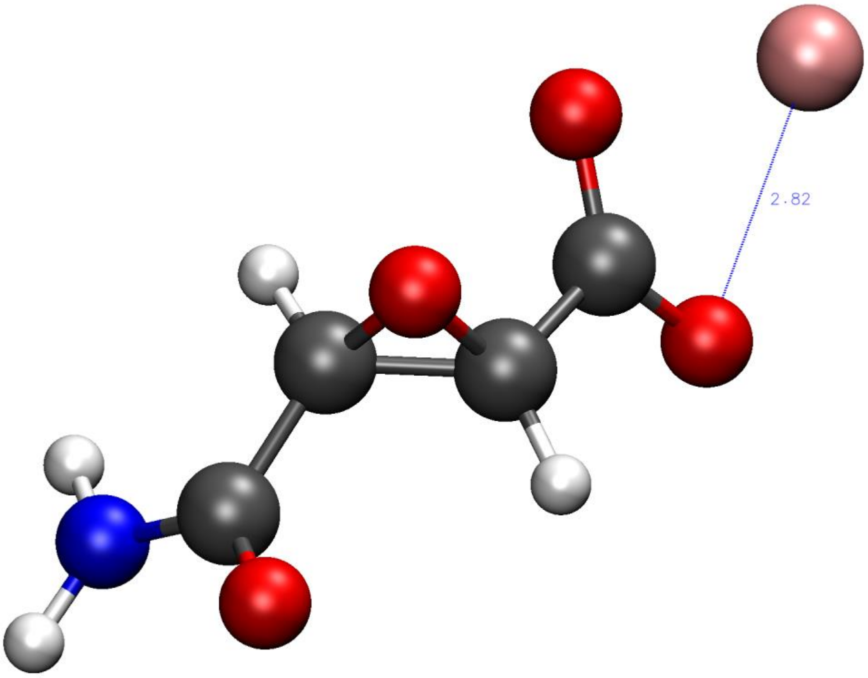


Figure S9: Geometry used in the model **S** (w K). The atomic radii used in the COSMO model are: K 2.223 Å; O 1.720 Å and C 2.000 Å. Therefore, it is ensured the solvation model does not separate the two entities from interacting with each other.

4. Difference in bond densities

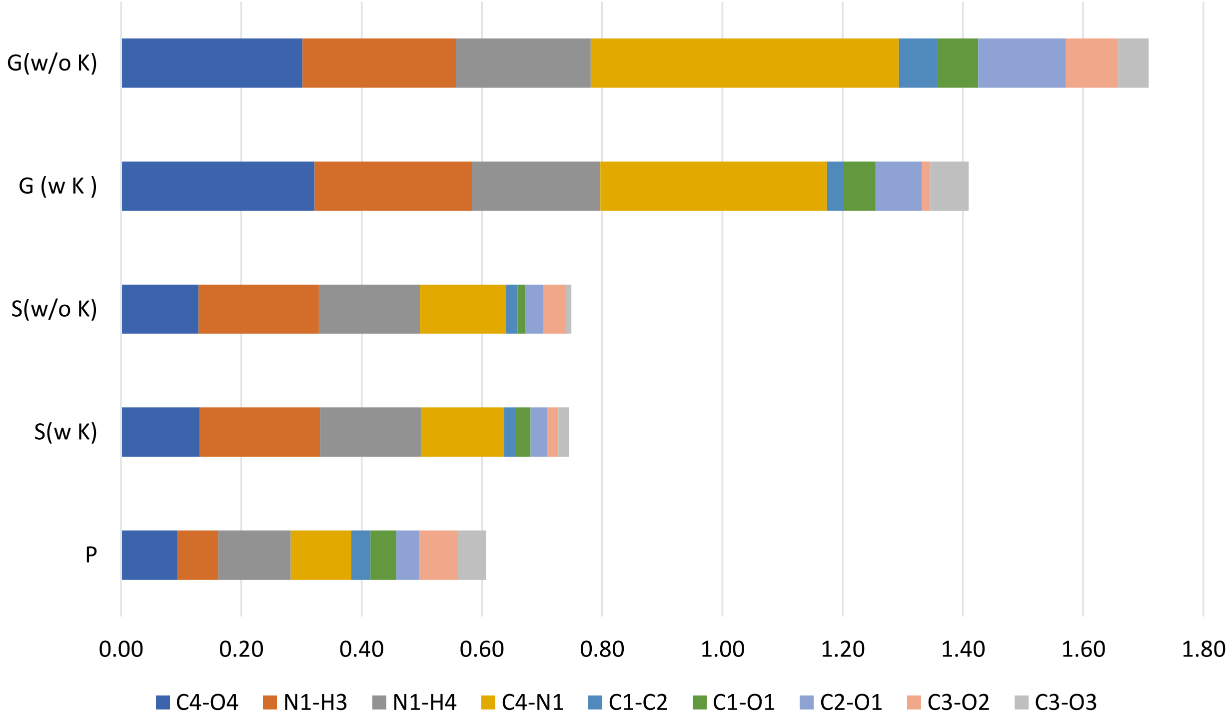


Figure S10: Sum of the number of electrons shifted (N_e in e) for all bond-centered difference density grids. The differences of the vacuum model (**G**) with or without K^+ counter-cation, the solvent model (**S**) with or without K^+ counter-cation and the enzyme model (**P**) are always calculated with respect to the crystal model (**C**).

D.2 Supplementary Material for *Publication:* Sila-Ibuprofen

This supplementary material was completely made by me except for chapter 7, which was prepared by the group of Prof. Dringen and the NMR spectra prepared by Daniel Duvinage and Pim Puylaert. All other figures, tables and text were made by me.

Supporting Information.

Sila-Ibuprofen

Florian Kleemiss, Aileen Justies, Daniel Duvinage, Patrick Watermann, Eric Ehrke, Kunihisa Sugimoto, Malte Fugel, Lorraine A. Malaspina, Anneke Dittmer, Torsten Kleemiss, Pim Puylaert, Nelly R. King, Anne Staubitz, Thomas M. Tzschentke, Ralf Dringen, Simon Grabowsky, Jens Beckmann

Contents of the Supporting Information

This PDF file includes:

1. Characterization
2. Quantum crystallography
3. Force-field development and molecular-dynamics simulations
4. Averaged non-covalent interaction index (aNCI)
5. Averaged interaction energies
6. Free energy perturbation (FEP) calculations
7. Toxicological investigation
8. Enzyme activity measurements to determine IC_{50} values
9. References to methods and software used in the bonding analysis
10. Captions of other Supporting Information
11. References for Supporting Information

Other Supporting Information for this manuscript include the following:

Movie S1
Data S1 to S4

1.) Characterization

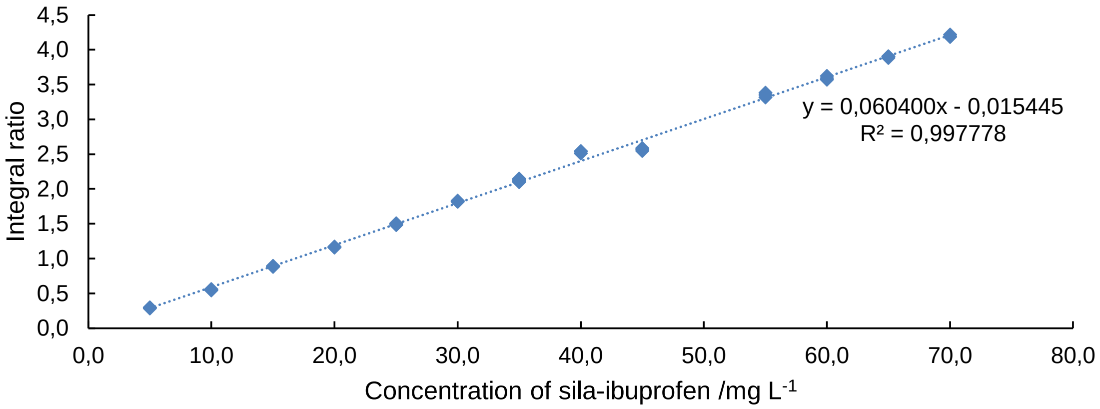


Fig. S1. Calibration curve of the HPLC/UV experiment for the determination of solubility of sila-ibuprofen (**2**) with ibuprofen (**1**) as internal standard. (Value at 50 *mg/L* was discarded due to machine failure)

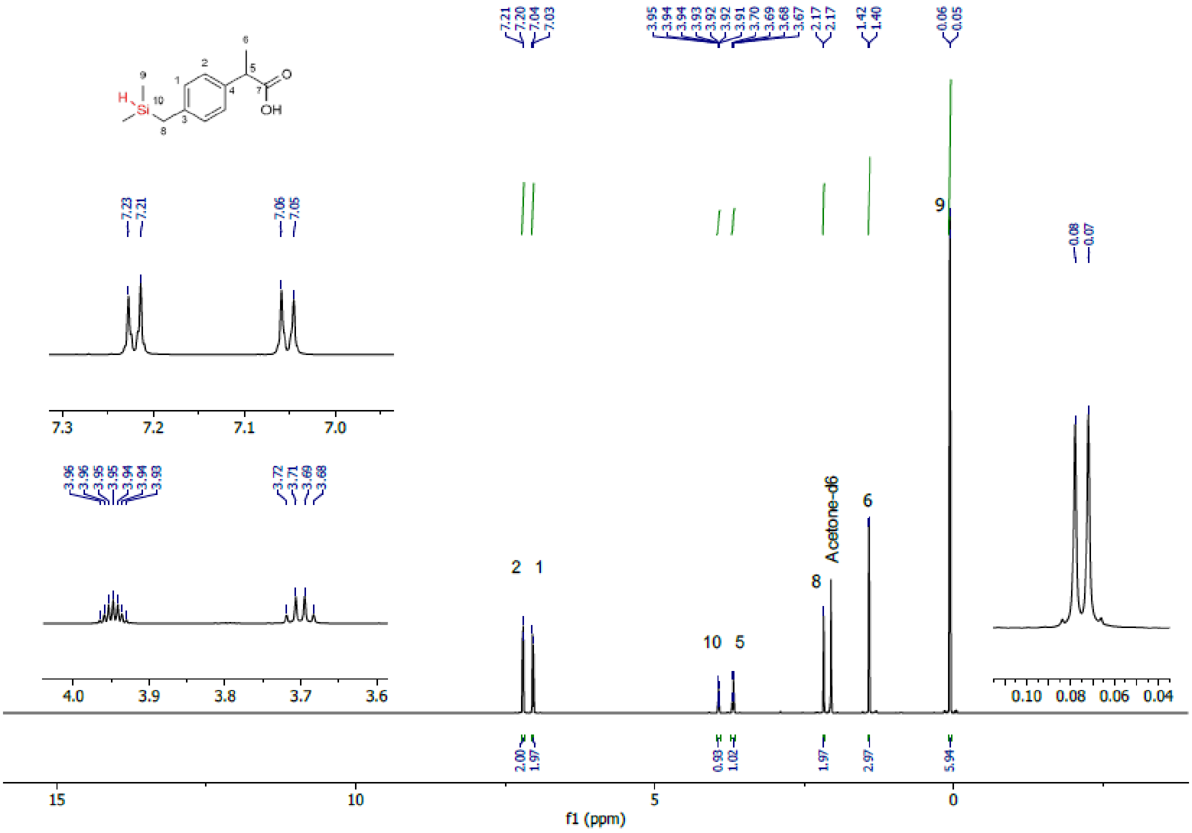


Fig. S2. ^{1}H-NMR spectrum (acetone-d6, 600 MHz) of sila-ibuprofen (**2**).

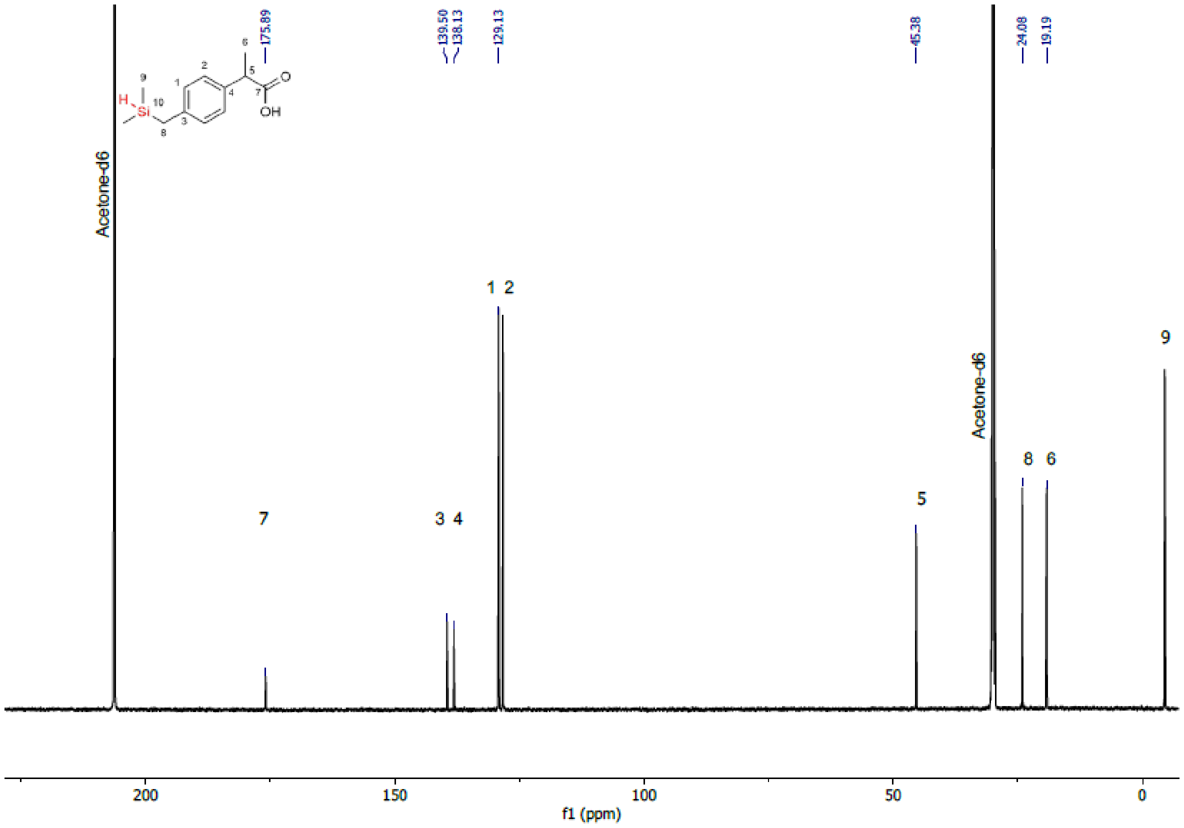


Fig. S3. ^{13}C-{^{1}H}-NMR spectrum (acetone-d6, 151 MHz) of sila-ibuprofen (**2**).

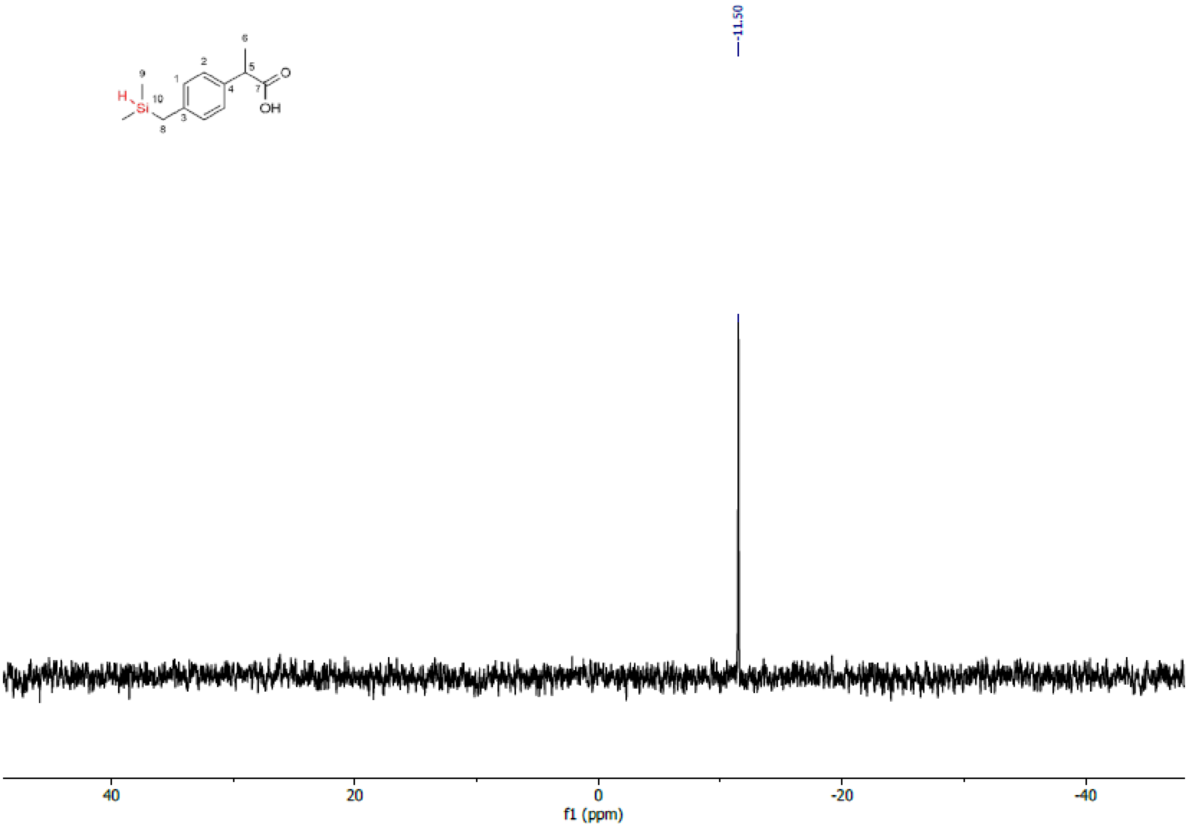


Fig. S4. ^{29}Si-{^{1}H}-NMR spectrum (acetone-d6, 119 MHz) of sila-ibuprofen (**2**).

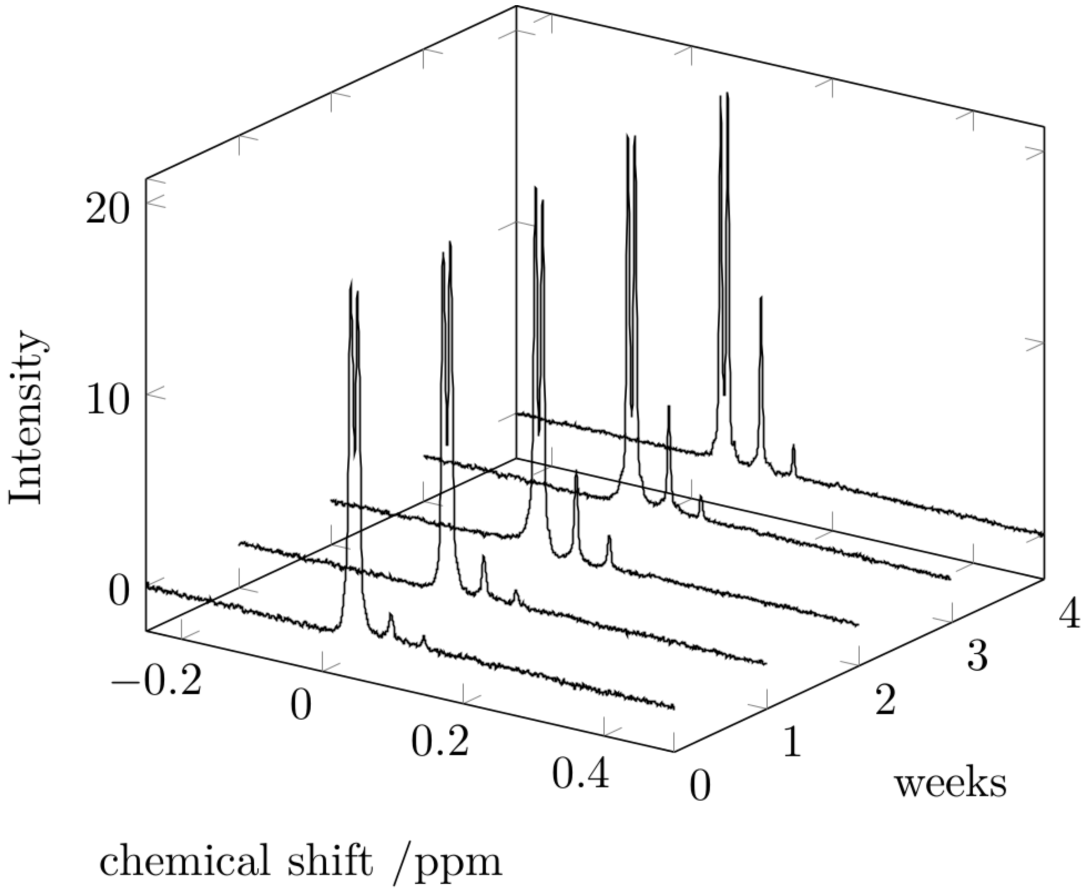


Fig. S5. ^{1}H-NMR spectra of the Me-Si group of sila-ibuprofen (**2**) in a 0.9% NaCl solution, kept at room temperature. Spectra were measured every 7 days. Decomposition is visible after one week.

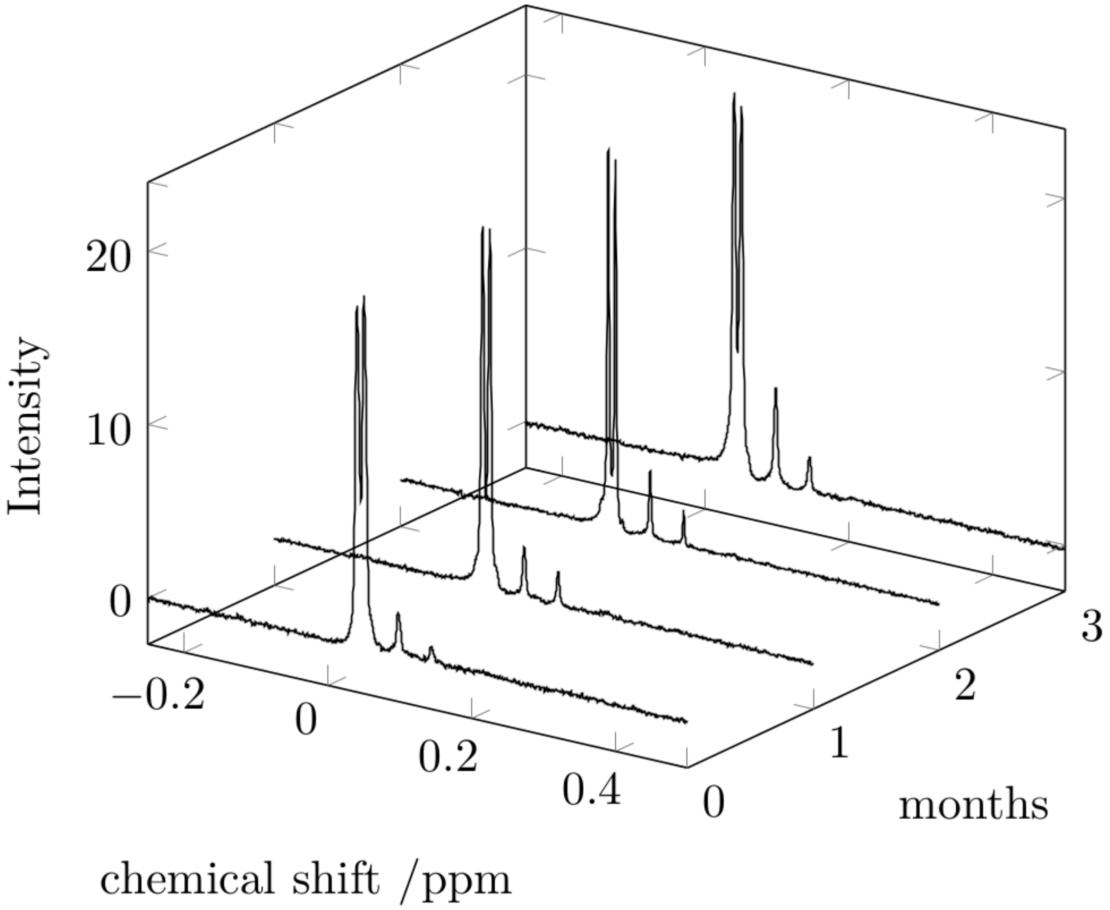


Fig. S6. ^{1}H-NMR spectra of the Me-Si group of sila-ibuprofen (**2**) in a 0.9% NaCl solution, kept at 4°C. Spectra were measured every 28 days. Decomposition is visible after more than a month.

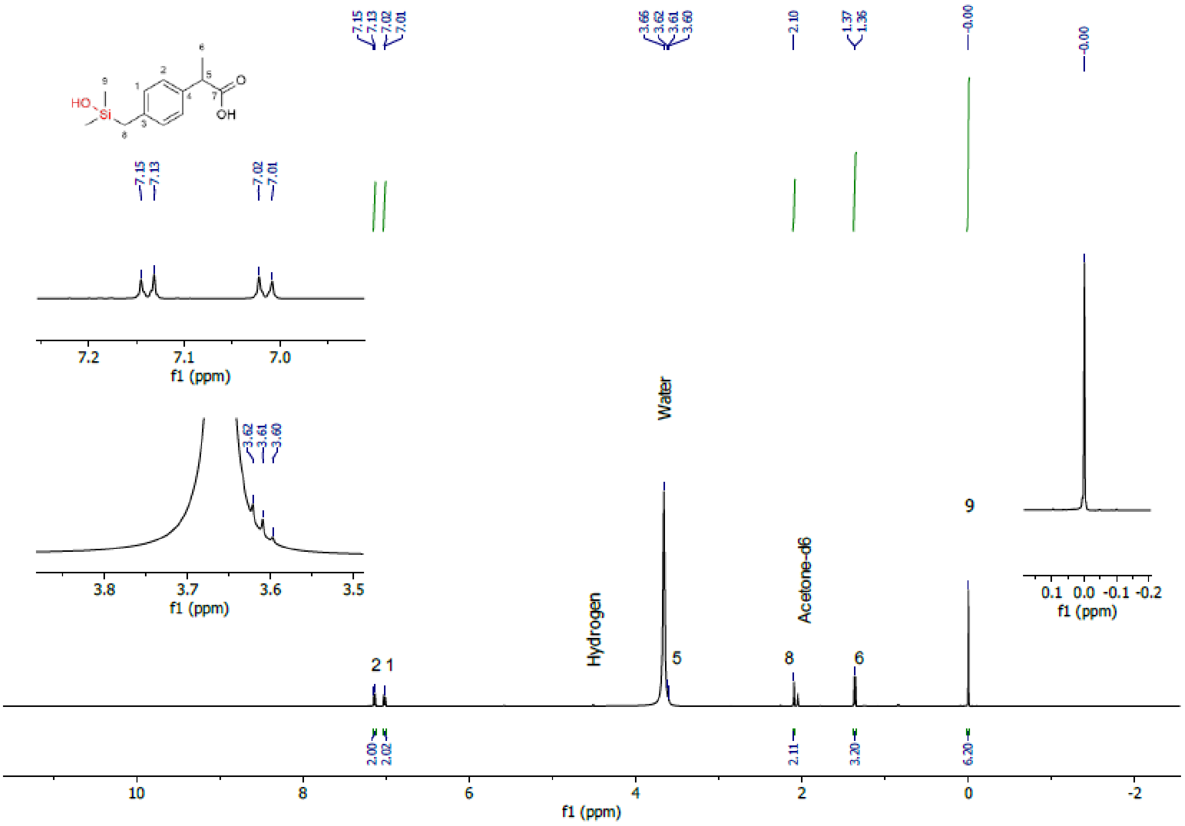


Fig. S7. ^{1}H-NMR spectrum (acetone-d6/water, 600 MHz) of hydroxysila-ibuprofen (**3**) directly from the reaction of sila-ibuprofen (**2**) with Pearlman's catalyst in an acetone-d6 – water mixture.

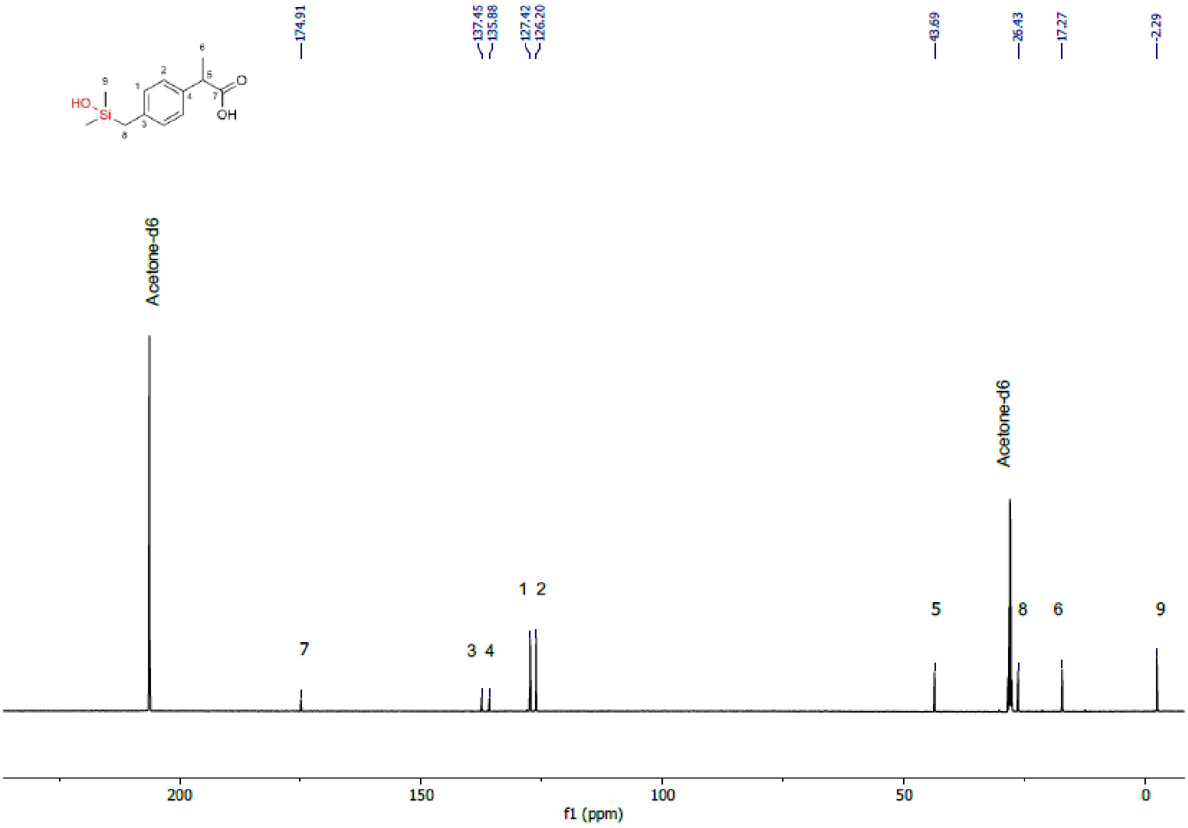


Fig. S8. ^{13}C-{^{1}H}-NMR spectrum (acetone-d6/water, 151 MHz) of hydroxysila-ibuprofen (**3**) directly from the reaction of sila-ibuprofen (**2**) with Pearlman's catalyst in an acetone-d6 – water mixture.

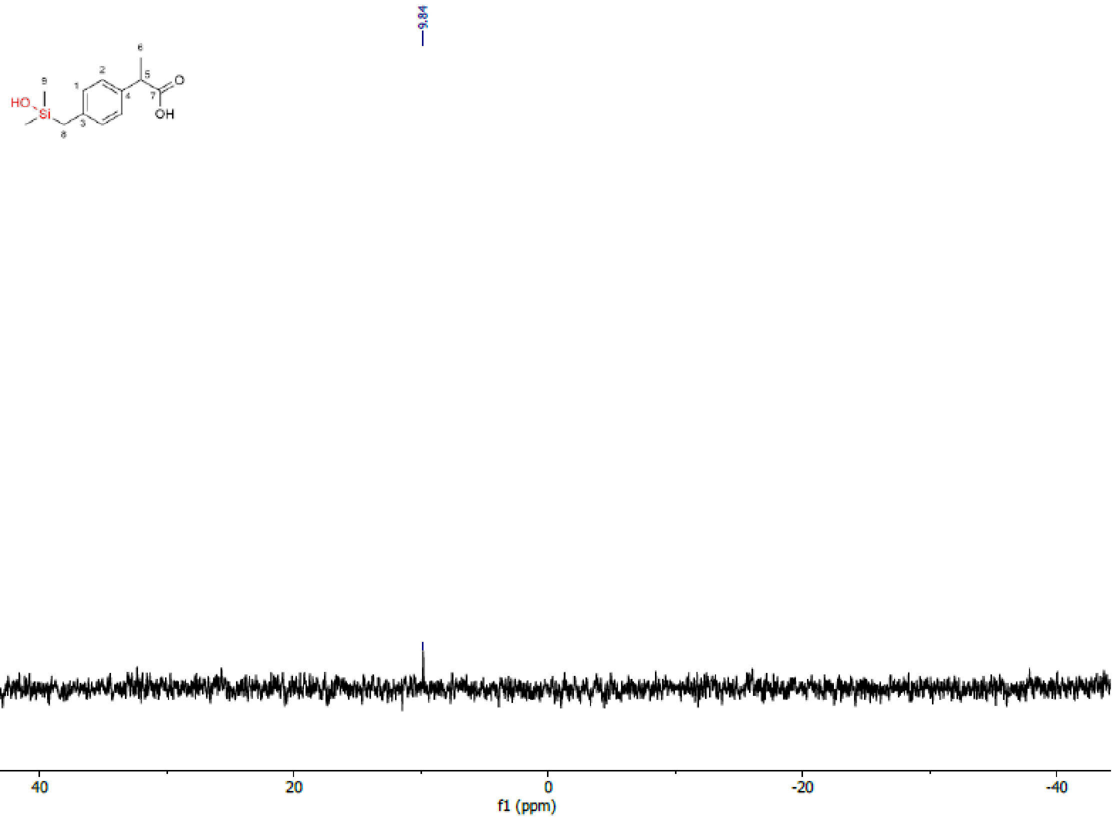


Fig. S9. ^{29}Si-{^{1}H}-NMR spectrum (acetone-d6/water, 119 MHz) of hydroxysila-ibuprofen (**3**) directly from the reaction of sila-ibuprofen (**2**) with Pearlman's catalyst in an acetone-d6 – water mixture.

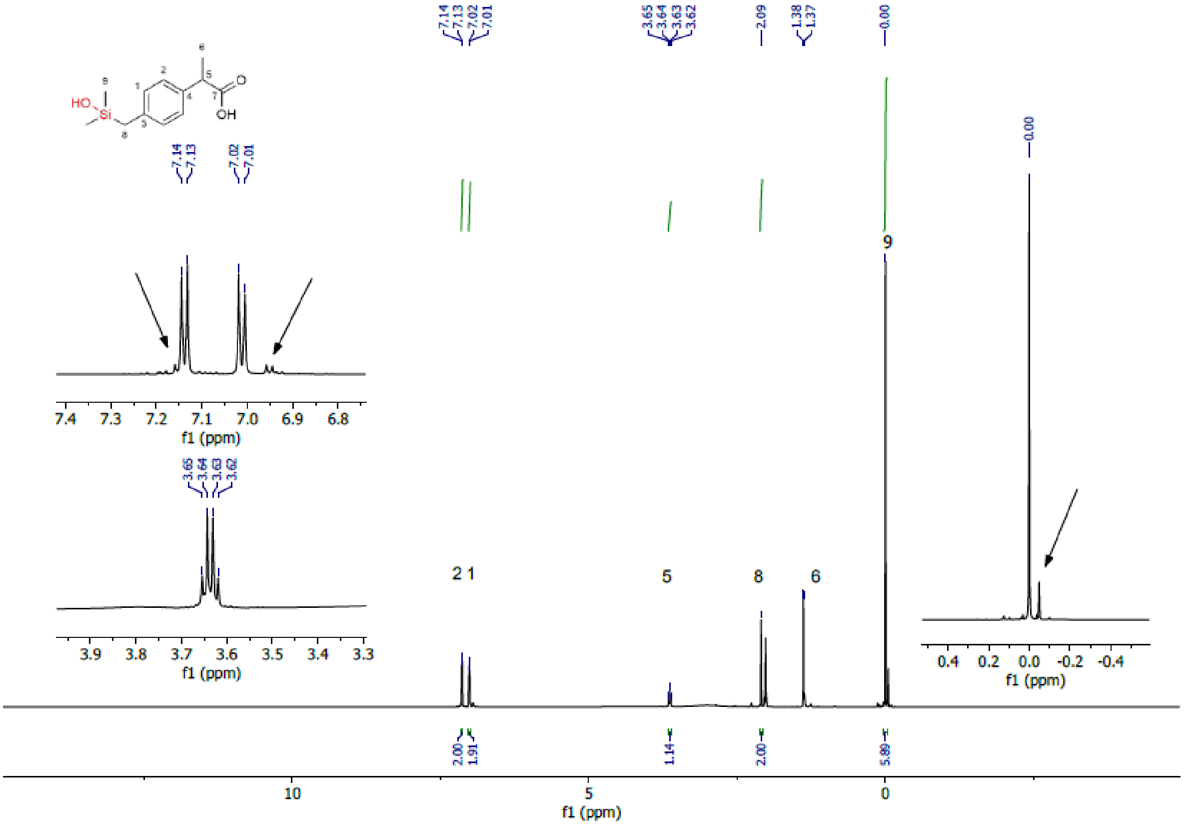


Fig. S10. ^{1}H-NMR spectrum (acetone-d6, 600 MHz) of hydroxysila-ibuprofen (**3**) after evaporation of the solvent showing condensation to disiloxane (**4**) at room temperature.

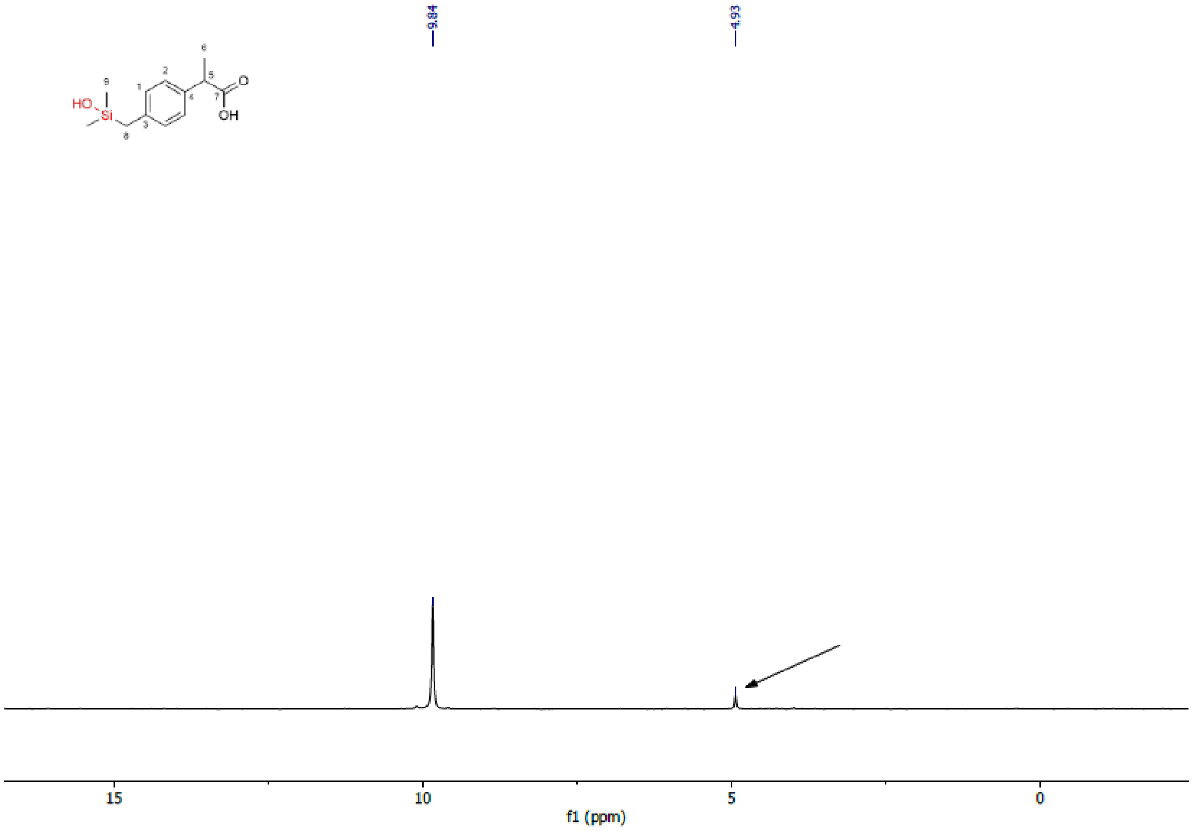


Fig. S11. ^{29}Si-{^{1}H}-NMR spectrum (acetone-d6, 119 MHz) of hydroxysila-ibuprofen (**3**) after evaporation of the solvent showing condensation to disiloxane (**4**).

2.) Quantum crystallography

Table S1. Refinement statistics of XWR for **1** and **2**.

Structure	Ibuprofen (**1**)	Sila-ibuprofen (**2**)
Space group	$P2_1/c$	$P2_1/c$
a /Å	14.465(3)	14.814(3)
b /Å	7.815(2)	7.972(2)
c /Å	10.435(2)	10.798(2)
β /°	99.66(3)	100.70(3)
V /Å^3	1162.9(4)	1253.0(4)
T /K	25	
Resolution /Å	0.45	
Wavelength /Å	0.3567	
R_{int}	0.0552	0.0557
Avg. redundancy	9.49	6.71
Completeness	1.00	1.00
Average I/σ	41.3	22.3
# of refln. measured	132378	105581
# of unique refln.	13367	14320
Criterium for observed refln.	$F_o^2 > 2\sigma(F_o^2)$	
# of observed refln.	10549	10780
Weighting scheme	$w = 1/\sigma(F_o)$	
Final λ	0.40	0.63
Final χ^2	1.7817	2.5273
Final R_1	0.0216	0.0267
Final wR_2	0.0285	0.0365
Max residual density /eÅ^{-3}	0.183	0.261
Min residual density /eÅ^{-3}	-0.139	-0.192
CCDC deposition number	1983628	1983627

Table S2. Full Table of Politzer parameters[14] in $e\text{Å}^{-1}$ or $e^2\text{Å}^{-2}$, respectively (V_+: Mean positive values, V_-: Mean negative Values, Π: average deviation from mean value on the surface, σ_+^2,σ_-^2 and σ_{Tot}^2 referring to variance of surface values for positive, negative and all values of the surface, respectively and ν with an upper limit of 0.250 referring to electrostatic balance in interactions of positive and negative regions) for 1 and 2. Calculated using CrystalExplorer.[32]

	Ibuprofen (**1**)	Sila-ibuprofen (**2**)
V_+	0.0217	0.0246
V_-	-0.0306	-0.0302
Π	0.0240	0.0263
σ_+^2	63548	78665
σ_-^2	36953	52630
σ_{Tot}^2	100501	131295
ν	0.2325	0.2401

3.) Force-field development and molecular-dynamics simulations

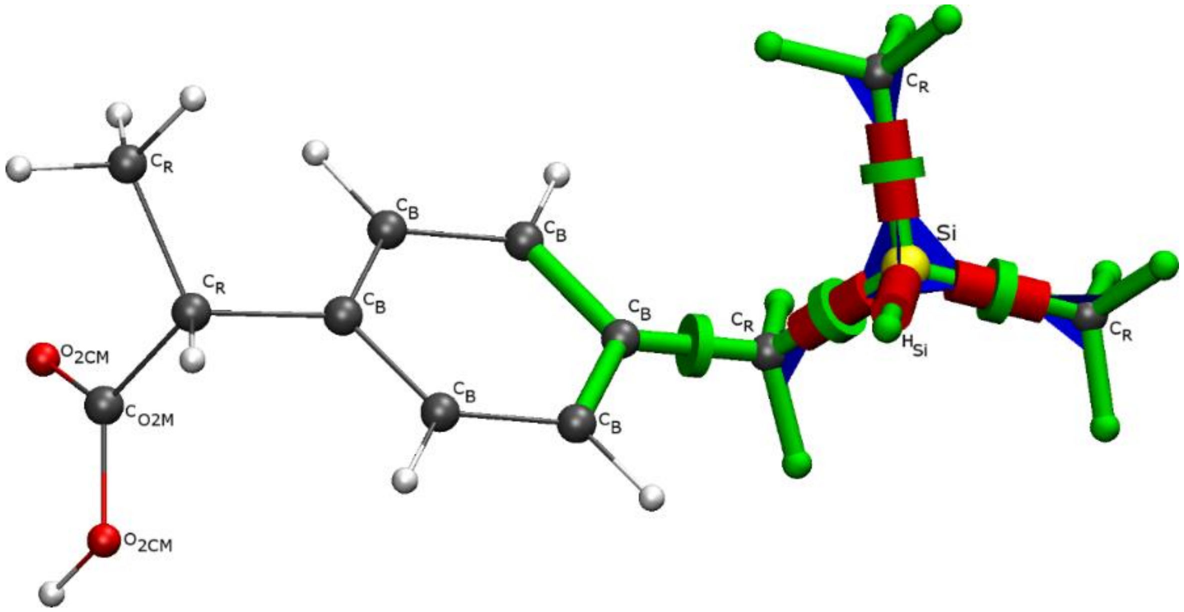


Fig. S12. Visualization of parameters needed to complete the forcefield of **2**. Red cylinders denote missing bond-, blue triangles missing angle- and green discs missing dihedral-parameters. All hydrogen atoms without labels are of type H_{CMM}.

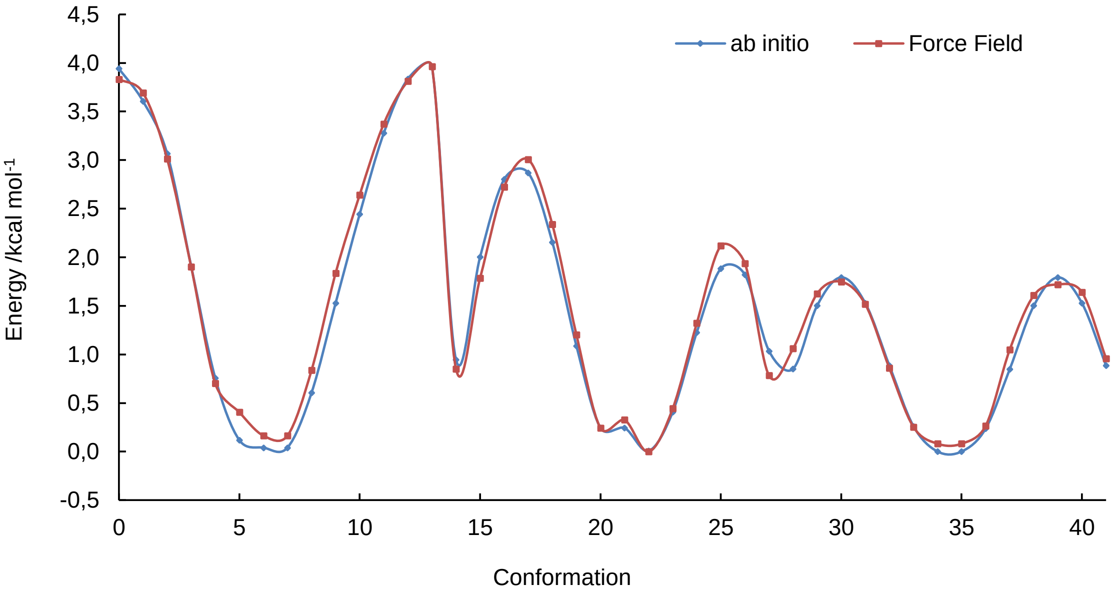


Fig. S13. Plot of the energy comparison between ab initio and force field for the scan around the torsion angles of type C_B-C_B-C_R-Si, C_B-C_R-Si-C_R and H_{CMM}-C_R-Si-H_{Si}.

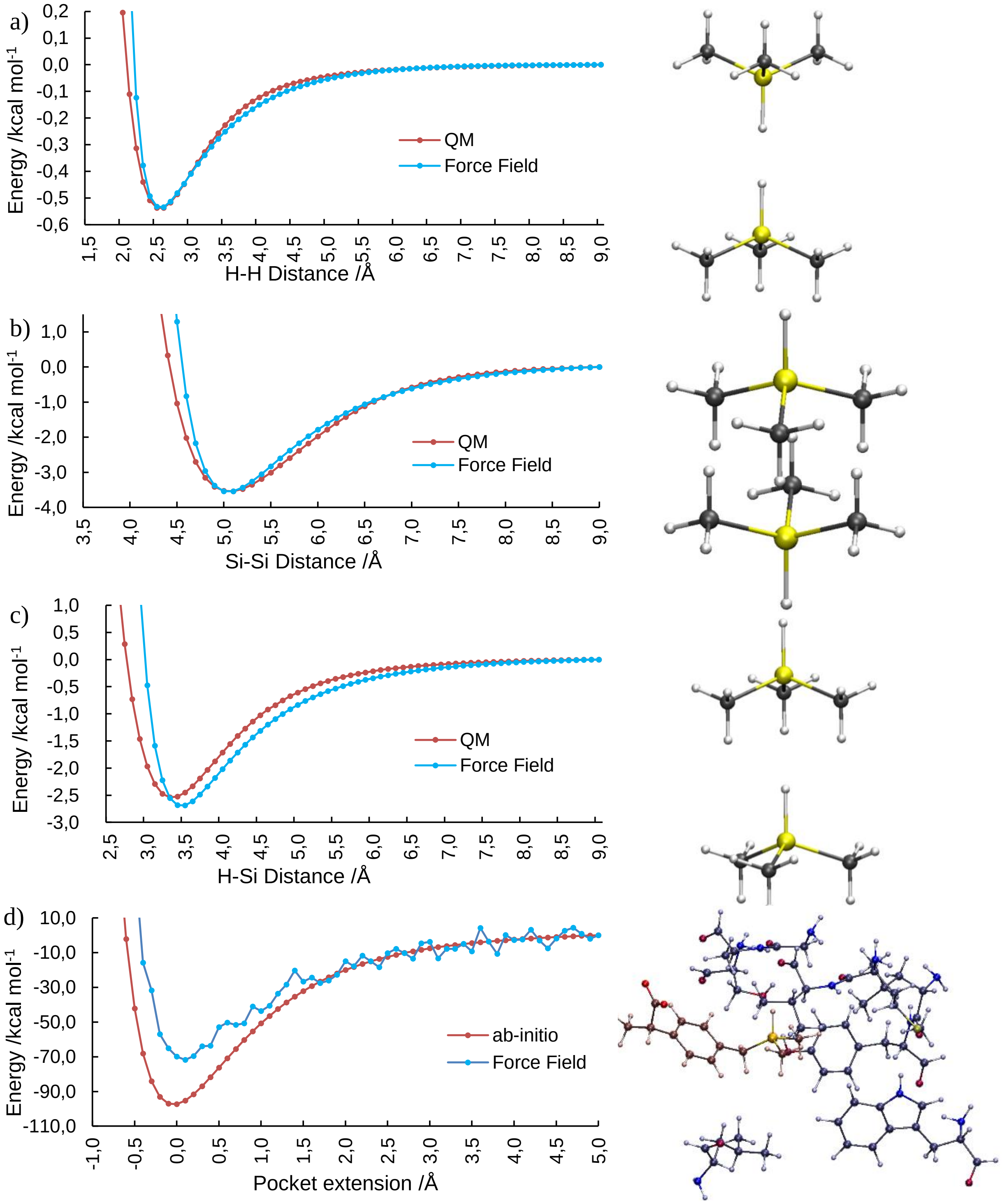


Fig. S14. Comparison of ab initio energies and force field energies for non-bonded interactions in various orientations in a model molecule (a-c) and the active site of the enzyme (d), with color coded sila-ibuprofen (2,red) and amino acids of COX-II (blue).

Table S3.
Parameters of the force field used for **1** and **2** in kcal mol^{-1} Å^{-2} and kcal mol^{-1} rad^{-2}, respectively. The parameters were derived and optimized using the ffTK program.[38]

Parameter (Atoms)	Force constant (K_b, K_θ)	Eq. geometry (b_0, θ_o)	Parameter (Atoms)	Force constant (K_χ)	Multiplicity (n)	Eq. geometry (δ)
Bond (C_R Si)	190.6411	1.8677	Dihedral (C_B C_B C_R Si)	0.442	2	180.00
Bond (Si H_{Si})	171.6976	1.4972	Dihedral (C_B C_R Si C_R)	0.331	3	0.00
Bond (C_{O2M} O_{2CM})[a]	702.1030	1.2610	Dihedral (C_B C_R Si H_{Si})	1.091	1	0.00
Bond (C_R C_R)[a]	306.4320	1.5080	Dihedral (C_R Si C_R H_{CMM})	0.000	3	0.00
Bond (C_R H_{CMM})[a]	342.9910	1.0930	Dihedral (H_{CMM} C_R Si H_{Si})	0.301	3	0.00
Bond (C_R C_B)[a]	356.7370	1.4860	Dihedral (C_B C_B C_B C_B)[a]	3.500	2	180.00
Bond (C_R C_{O2CM})[a]	275.6310	1.5100	Dihedral (C_B C_B C_B H_{CMM})[a]	3.500	2	180.00
Bond (C_B C_B)[a]	401.0680	1.3740	Dihedral (C_B C_R C_R H_{CMM})[a]	0.195	3	0.00
Bond (C_B H_{CMM})[a]	381.8530	1.0840	Dihedral (C_B C_R C_{O2M} O_{2CM})[a]	0.300	2	180.00
Angle (C_B C_R Si)	30.2479	105.4005	Dihedral (C_B C_B C_R C_R)[a]	0.225	2	180.00
Angle (C_R Si H_{Si})	22.5564	116.2359	Dihedral (C_B C_B C_R H_{CMM})[a]	-0.210	2	180.00
Angle (C_R Si C_R)	54.1973	114.2112	Dihedral (C_B C_B C_R H_{CMM})[a]	0.196	3	0.00
Angle (Si C_R H_{CMM})	14.9964	110.2184	Dihedral (C_B C_B C_R C_{O2M})[a]	0.100	3	0.00
Angle (C_B C_B C_B)[a]	48.1450	119.9770	Dihedral (C_B C_B C_B C_R)[a]	3.500	2	180.00
Angle (C_B C_B C_R)[a]	57.7880	120.4190	Dihedral (C_B C_R C_R C_R)[a]	0.150	3	0.00
Angle (C_B C_B H_{CMM})[a]	40.5170	120.5710	Dihedral (C_R C_R C_{O2M} O_{2CM})[a]	0.631	2	180.00
Angle (C_R C_R H_{CMM})[a]	45.7700	110.5490	Dihedral (C_R C_B C_B H_{CMM})[a]	3.500	2	180.00
Angle (H_{CMM} C_R H_{CMM})[a]	37.1340	108.8360	Dihedral (H_{CMM} C_R C_R H_{CMM})[a]	0.142	1	0.00
Angle (C_B C_R C_R)[a]	54.4060	108.6170	Dihedral (H_{CMM} C_R C_R H_{CMM})[a]	-0.693	2	180.00
Angle (C_B C_R H_{CMM})[a]	45.1220	109.4910	Dihedral (H_{CMM} C_R C_R H_{CMM})[a]	0.157	3	0.00
Angle (C_B C_R C_{O2M})[a]	71.9660	109.5000	Dihedral (H_{CMM} C_R C_{O2M} O_{2CM})[a]	-0.053	3	0.00
Angle (C_R C_R C_{O2M})[a]	23.7490	98.4220	Dihedral (C_R C_R C_R H_{CMM})[a]	0.320	1	0.00
Angle (H_{CMM} C_R C_{O2M})[a]	37.7820	108.9040	Dihedral (C_R C_R C_R H_{CMM})[a]	-0.315	2	180.00
Angle (C_R C_R C_R)[a]	61.2430	109.6080	Dihedral (C_R C_R C_R H_{CMM})[a]	0.132	3	0.00
Angle (C_R C_{O2M} O_{2CM})[a]	87.0070	114.6890	Dihedral (H_{CMM} C_B C_B H_{CMM})[a]	3.500	2	180.00
Angle (O_{2CM} C_{O2M} O_{2CM})[a]	84.9910	130.6000	Dihedral (C_{O2M} C_R C_R H_{CMM})[a]	-0.070	3	0.00
			Improper (C_B C_B C_B C_R)[a]	2.879	-	-
			Improper (C_B C_B C_B H_{CMM})[a]	1.079	-	-
			Improper (C_B C_R C_B C_B)[a]	2.879	-	-
			Improper (C_{O2M} O_{2CM} C_R O_{2CM})[a]	12.810	-	-

[a] These parameters were taken from swissparam.[28]

Table S4.
Lennard-Jones interaction parameters of the force field.

Atom Type	ε /kcal mol^{-1}	R$_{min}$ /Å
$C_{B}{}^{a}$	-0.0700	1.9924
$C_{R}{}^{a}$	-0.0550	2.1750
$H_{CMM}{}^{a}$	-0.0220	1.3200
$C_{O2M}{}^{a}$	-0.0700	2.0000
$O_{2CM}{}^{a}$	-0.1200	1.7000
H_{Si}	-0.0152	1.5210
Si	-0.5650	2.3800

[a]These parameters were taken from swissparam.[28]

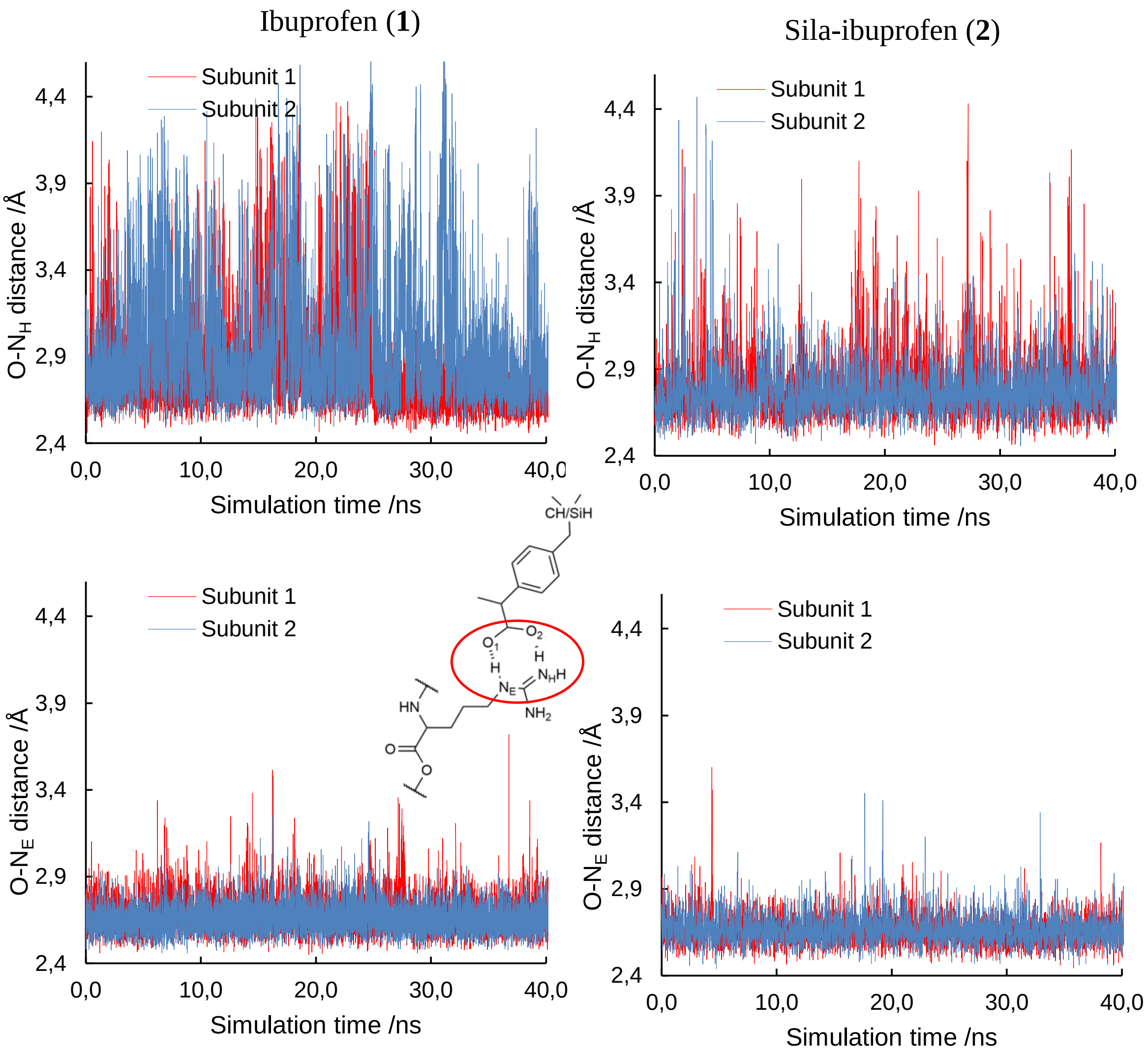
Table S5.
Charges of atoms in the force field used for **1** and **2**.

Atom Type	Ibuprofen[a] (**1**)	Sila-ibuprofen (**2**)
O_{2CM}	-0.9000	-0.9000
$C_{B,q}$	-0.1435	-0.1435
$C_{B,t}$	-0.1500	-0.1500
C_{02M}	0.9060	0.9060
$C_{R,CH2}$	0.1435	0.1480
$C_{R,CH}$	0.0375	0.0375
$C_{R,CH3\text{-}C/Si}$	0.0000	0.0000
H_{CR}	0.0000	0.0000
H_{CB}	0.1500	0.1500
H_{CH}/H_{Si}	0.0000	-0.1470
C_{H}/Si	0.0000	0.1430

[a]These parameters were taken from swissparam.[28]

The most significant difference in the description between **1** and **2** is the assigned charge to the tertiary position where the carbon-silicon-switch was performed. While **1** was described without charges on neither carbon nor hydrogen, as usual in CHARMM force fields,[29] the significant pronunciation of charges observed in the XWR (compare Figure 2) made clear an explicit charge was necessary for **2**.

The parameters in Tables S3 to S5 show not only close resemblance to the quantum mechanically derived energies (compare Figures S5 and S6), but also are in similar orders of magnitude as other parameters used in the CHARMM force field, pointing towards reasonable results of the optimization process.



Ibuprofen (**1**)

Sila-ibuprofen (**2**)

Fig. S15. Plot of arginine-ibuprofen (**1**,left) and -sila-ibuprofen (**2**,right) distances for O_1-N_E and O_2-N_H atoms (in force-field nomenclature, structure shown in figure). There are two subunits in COX-II, which are plotted by color-code (red and blue).

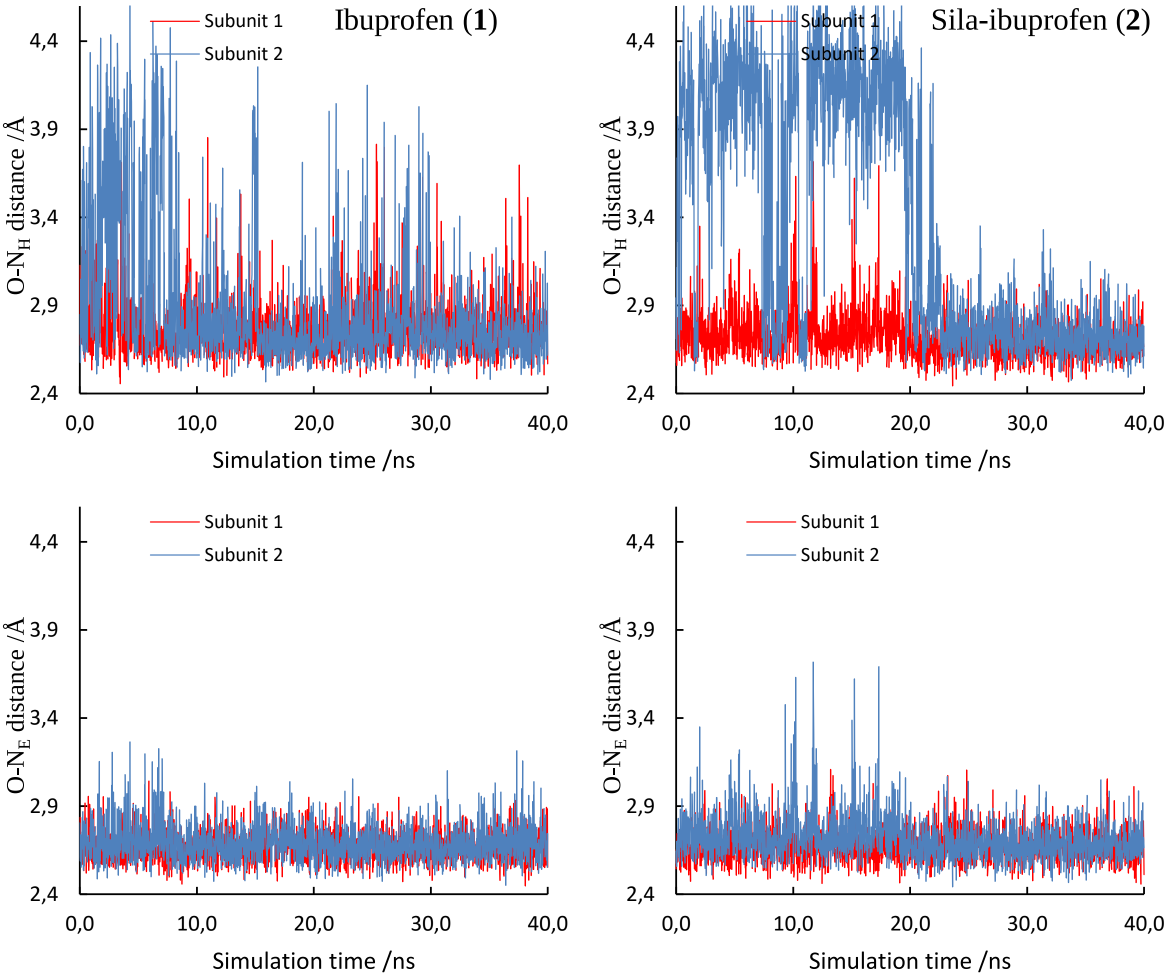


Fig. S16. Plot of arginine-ibuprofen (1,left) and -sila-ibuprofen (2,right) distances for O$_1$-N$_E$ and O$_2$-N$_H$ atoms (in force-field nomenclature, structure shown in Figure S7). There are two subunits in COX-I, which are plotted by color-code (red and blue).

4.) Averaged non-covalent interaction index (aNCI)

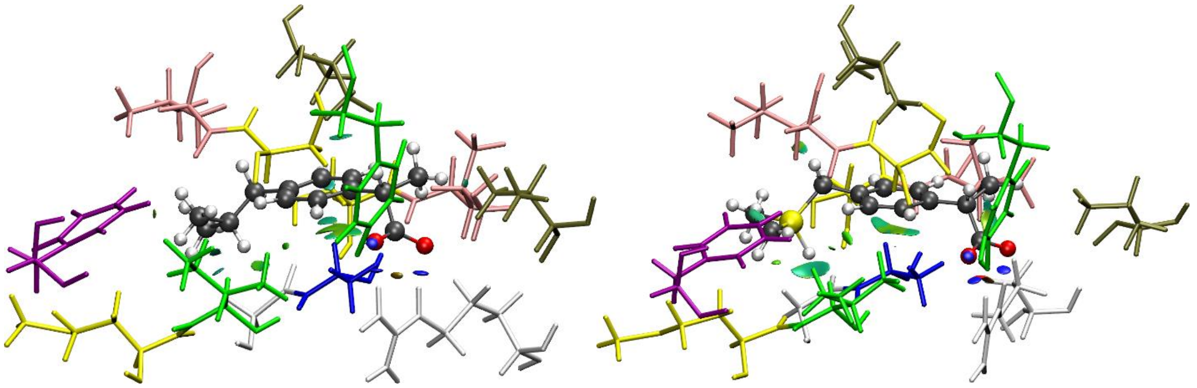


Fig. S17. Visualization of residues important for close interactions inside the active site of COX-I after MD of ibuprofen (**1**,left) and sila-ibuprofen (**2**,right) (color code in Figure 2). Visualization of the aNCI, color code: green = weak dispersion interactions, blue = stronger electrostatic interactions, orange = repulsive interactions.

5.) Averaged interaction energies

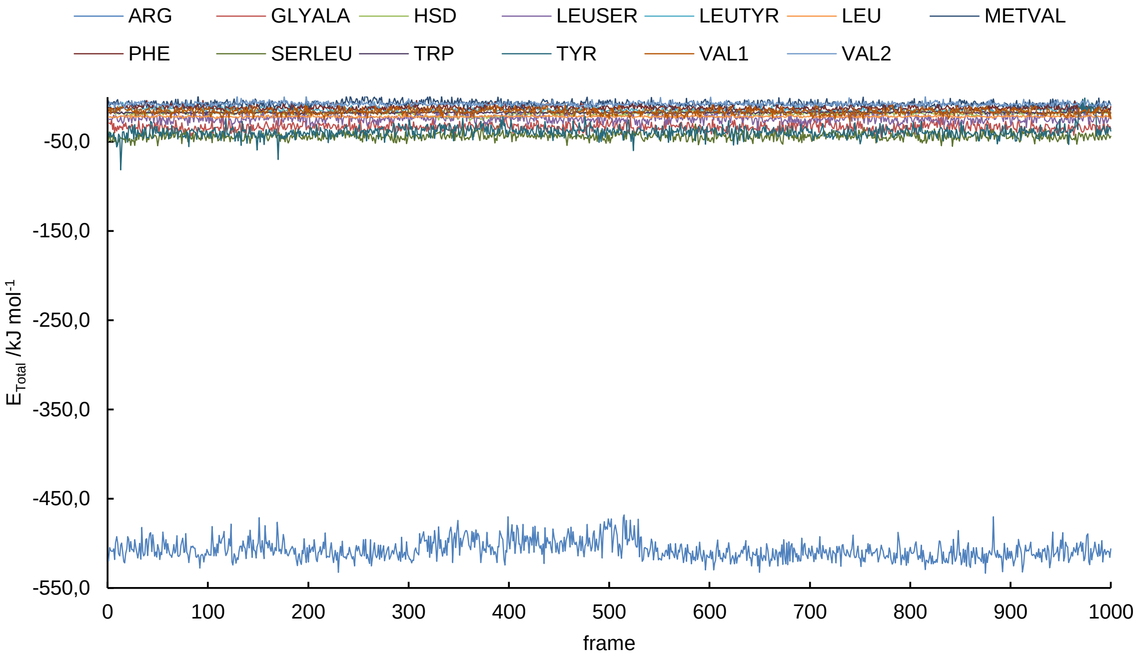


Fig. S18. Plot of total interaction energy over time for sila-ibuprofen (2) in COX-II with 13 residues, which had atoms in a 4 Å radius around atoms of 2.

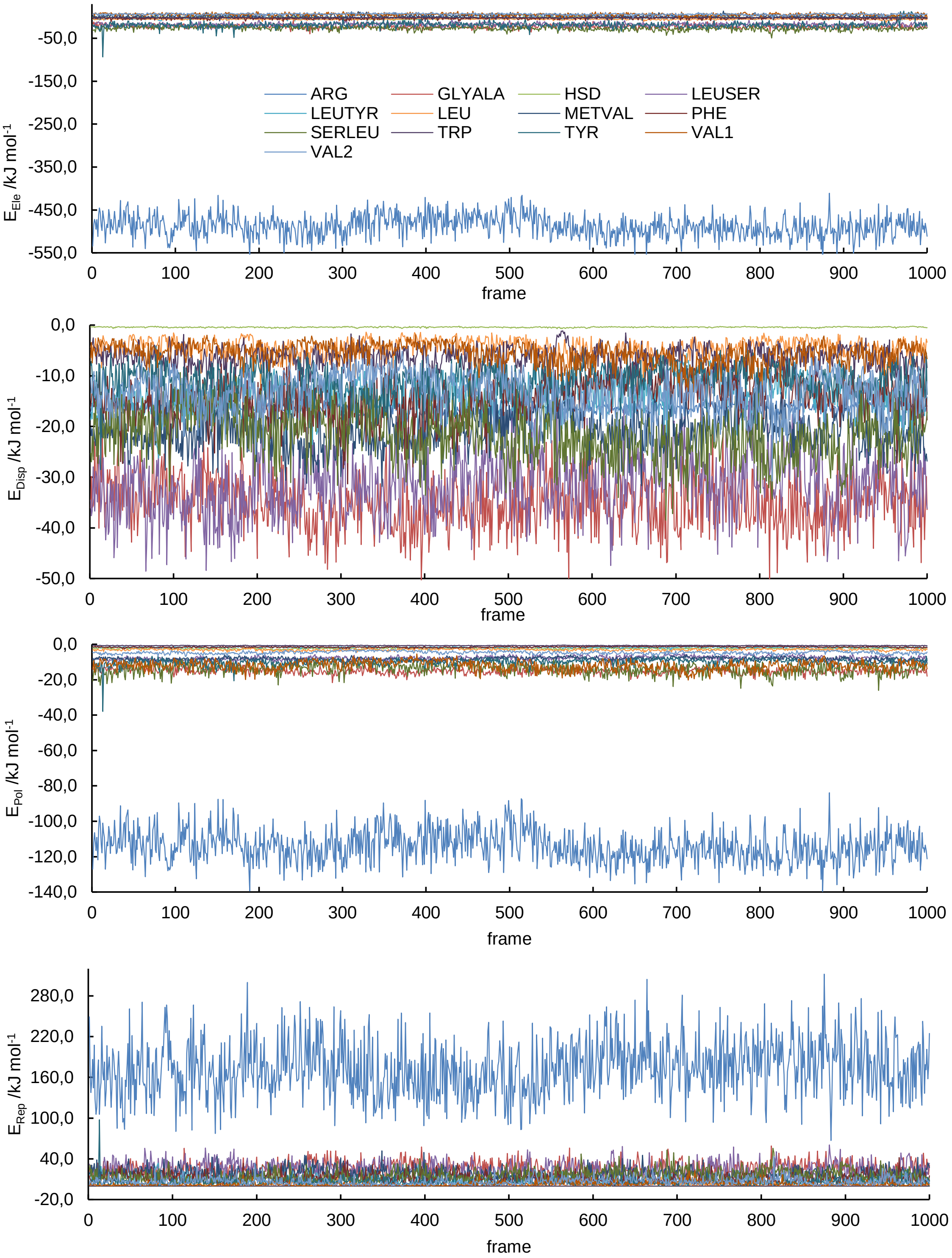


Fig. S19. Plot of interaction energy contributions over time for sila-ibuprofen (2) in COX-II with 13 residues, which had atoms in a 4 Å radius around atoms of 2: electrostatic (top), dispersion (mid, upper), polarization (mid, lower) and exchange repulsion (bottom).

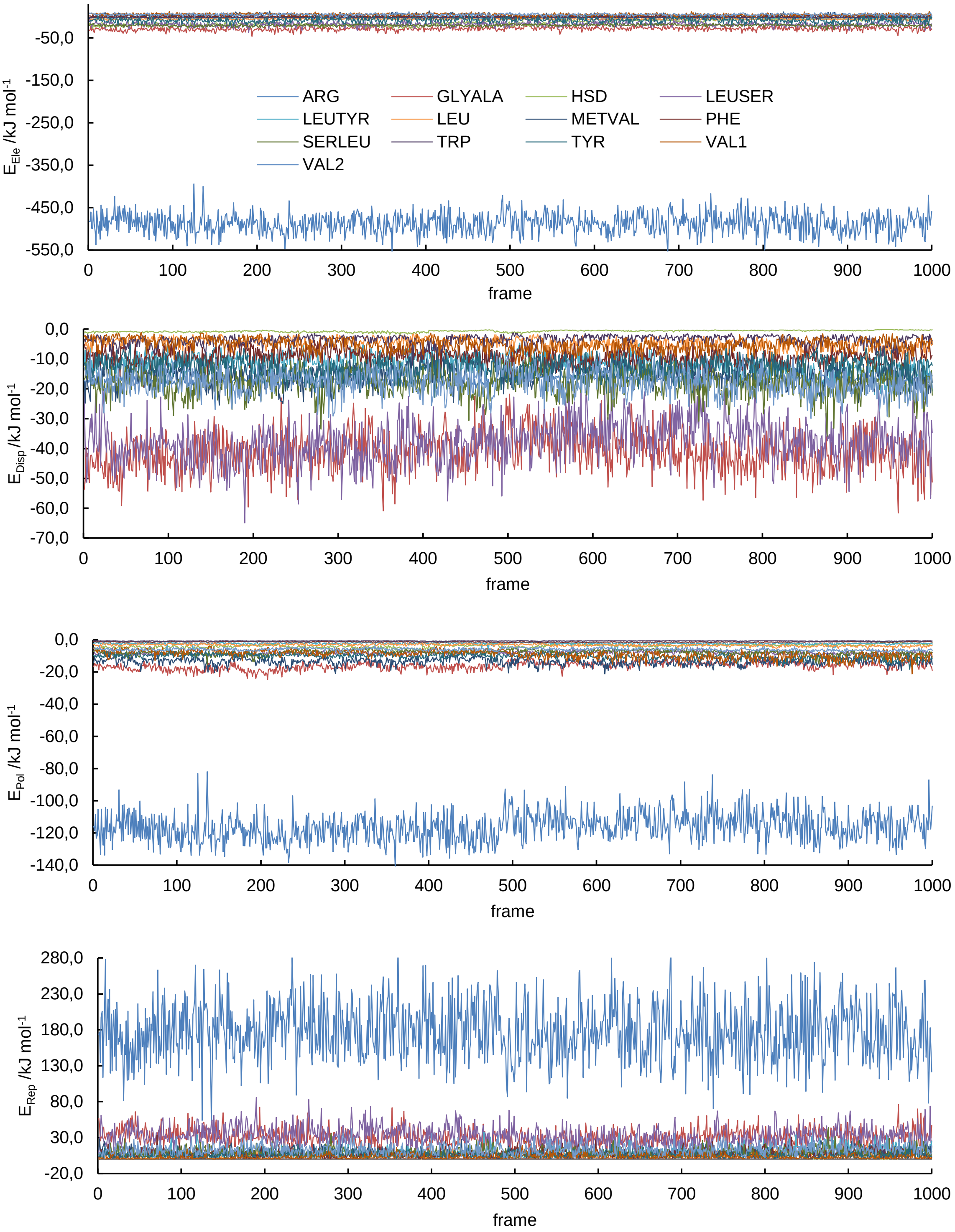


Fig. S20. Plot of interaction energy contribution over time for ibuprofen (1) in COX-II with 13 residues, which had atoms in a 4 Å radius around atoms of 1: electrostatic (top), dispersion (mid, upper), polarization (mid, lower) and exchange repulsion (bottom).

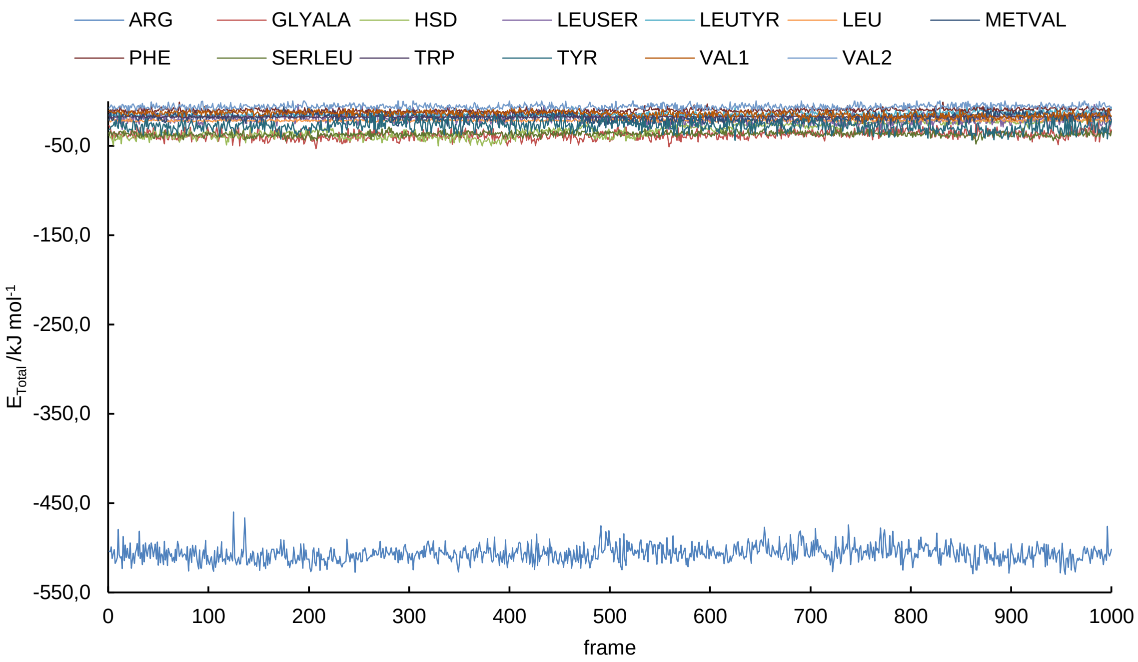


Fig. S21. Plot of total interaction energy over time for ibuprofen (1) in COX-II with 13 residues, which had atoms in a 4 Å radius around atoms of 1.

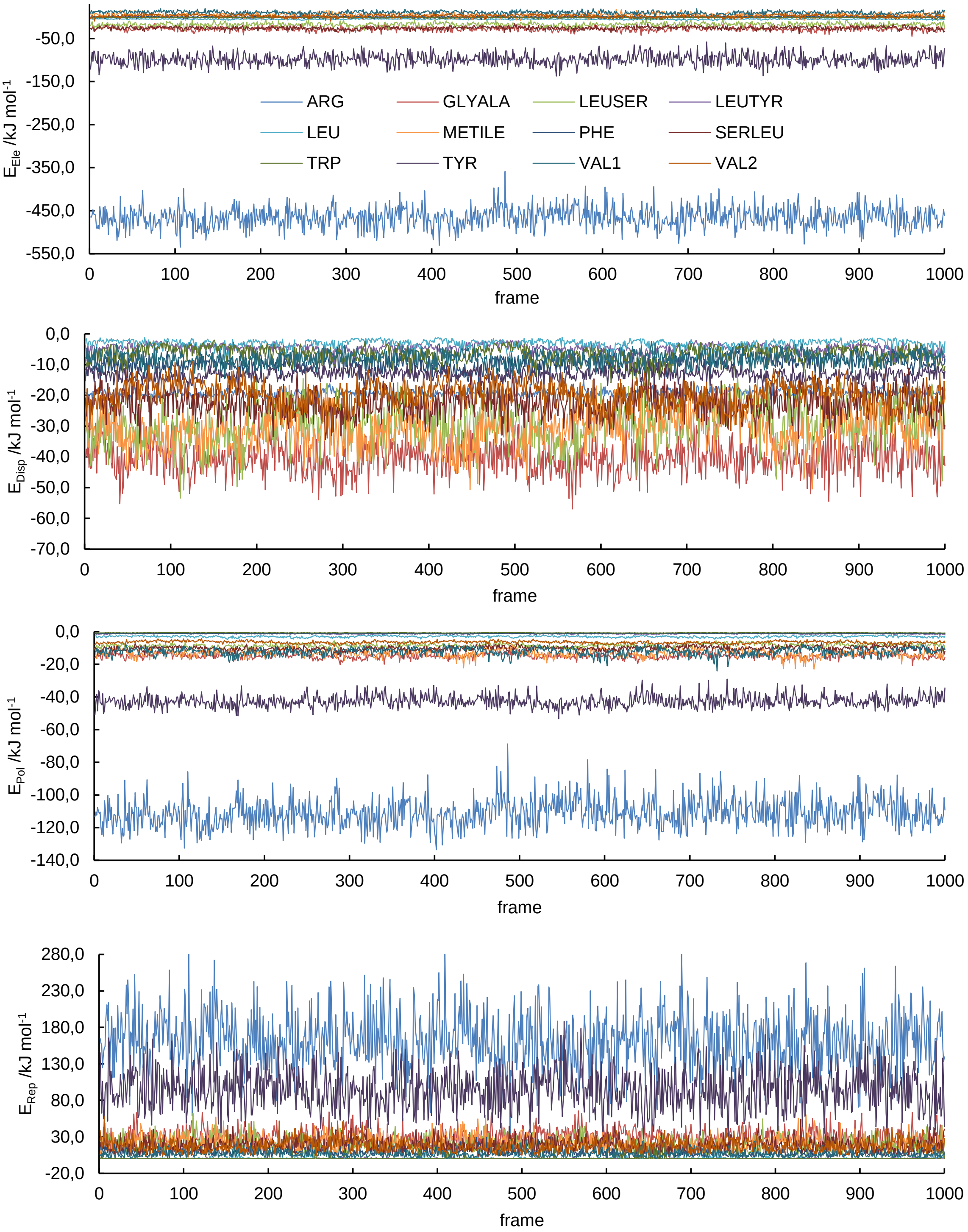


Fig. S22. Plot of interaction energy contributions over time for sila-ibuprofen (**2**) in COX-I with 12 residues, which had atoms in a 4 Å radius around atoms of **2**: electrostatic (top), dispersion (mid, upper), polarization (mid, lower) and exchange repulsion (bottom).

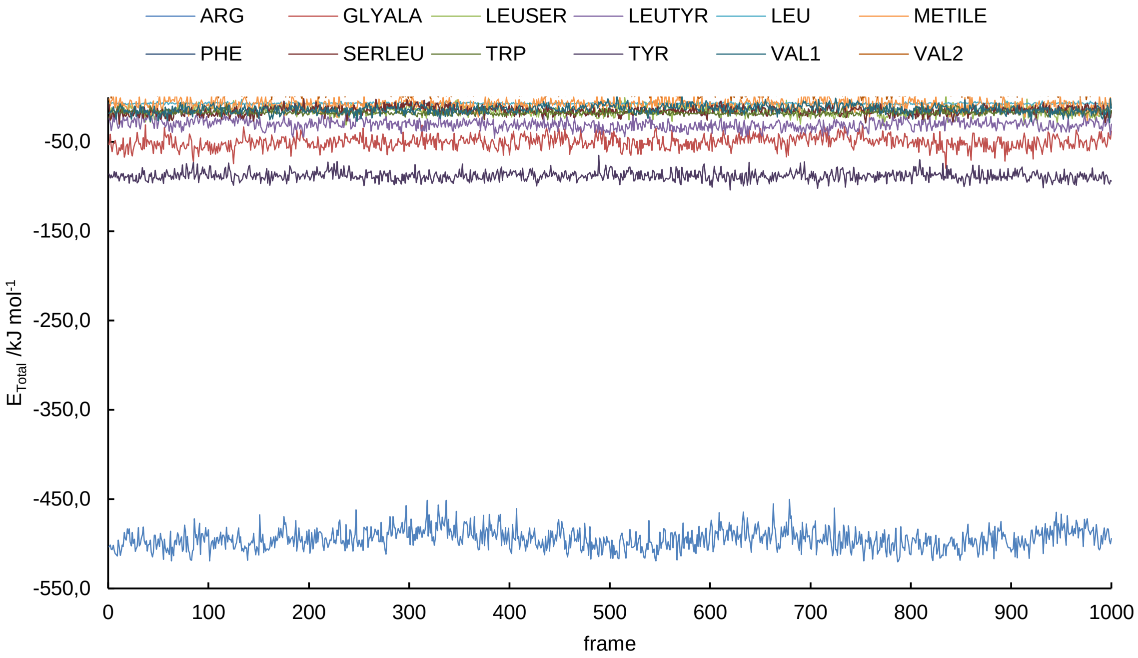


Fig. S23. Plot of total interaction energy over time for sila-ibuprofen (2) in COX-I with 12 residues, which had atoms in a 4 Å radius around atoms of 2.

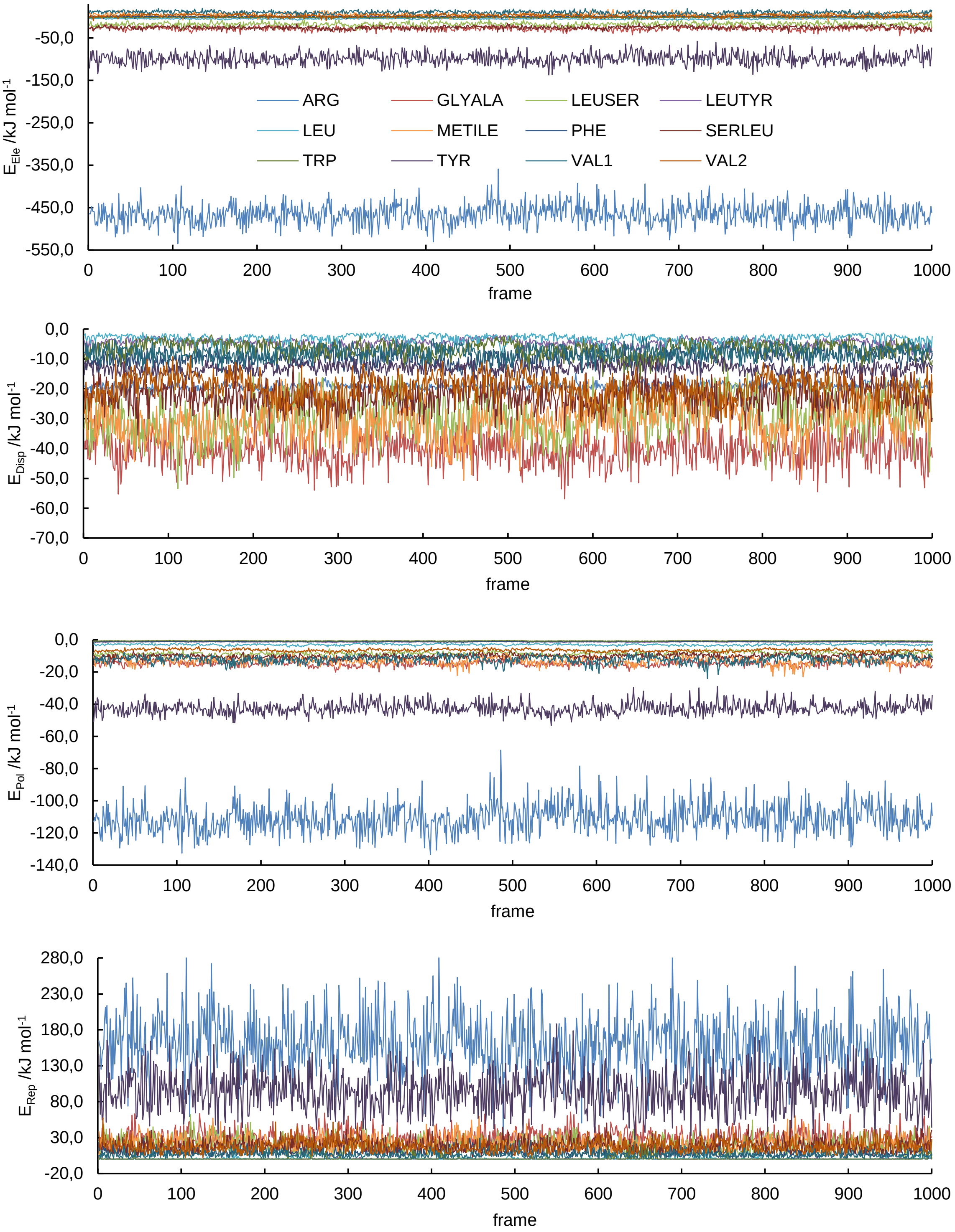


Fig. S24. Plot of interaction energy contributions over time for ibuprofen (**1**) in COX-I with 12 residues, which had atoms in a 4 Å radius around atoms of **1**: electrostatic (top), dispersion (mid, upper), polarization (mid, lower) and exchange repulsion (bottom).

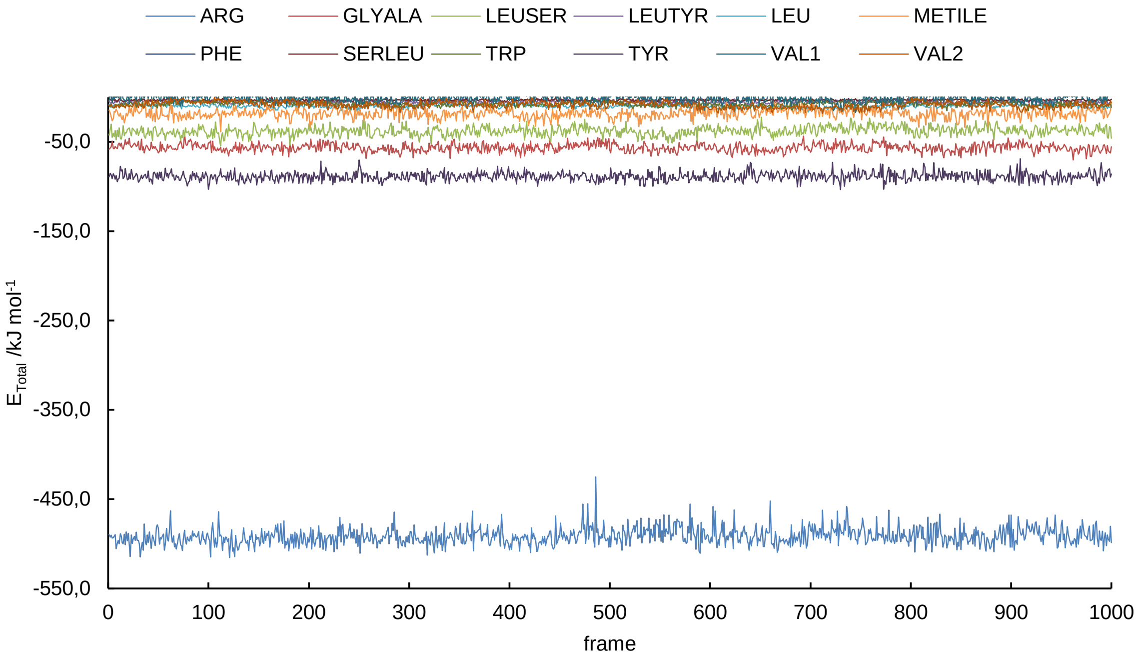


Fig. S25. Plot of total interaction energy over time for ibuprofen (1) in COX-I with 12 residues, which had atoms in a 4 Å radius around atoms of 1.

Table S6.
Full Table of averaged interaction energies in COX-II with estimated standard deviation (ESD) from averaging over trajectories in kJ/mol. Calculated using *tonto*, backend of *CrystalExplorer*.[32]

	Ibuprofen (**1**)					Sila-ibuprofen (**2**)				
Residue	Ele	Disp	Pol	Rep	Total	Ele	Disp	Pol	Rep	Total
Arg	-487(24)	-17.0(10)	-116(9)	178(40)	-508(9)	-487(25)	-16.3(9)	-114(9)	173(41)	-508(11)
GlyAla	-28(4)	-41(6)	-16(2)	31.0(11)	-38(5)	-22(4)	-35(5)	-15.1(15)	26(9)	-34(4)
Hsd	-14(7)	-0.7(3)	-3.8(10)	0(0)	-32(8)	-1.6(18)	-0.41(8)	-2.4(3)	0(0)	-19(2)
LeuSer	-19(5)	-38(7)	-8.2(9)	33(13)	-20(5)	-18(4)	-31(6)	-7.7(11)	25(1)	-24(4)
LeuTyr	-3(2)	-12(3)	-2.4(6)	9(6)	-15(2)	-2.9(17)	-14(3)	-1.7(3)	9(6)	-14(2)
Leu	-5.8(13)	-7(2)	-3.3(6)	3(3)	-21.7(10)	-5.4(11)	-4.4(17)	-3.1(5)	2(2)	-21.9(8)
MetVal	2(4)	-15(3)	-14(2)	8(5)	-18(4)	2(3)	-21(4)	-8.3(10)	18(8)	-8(3)
Phe	0.8(14)	-9(2)	-1.36(15)	8(5)	-10.2(19)	-3(2)	-14(4)	-1.5(3)	12(7)	-12(2)
SerLeu	-20(2)	-18(4)	-8.1(16)	9(6)	-37(3)	-27(4)	-22(5)	-14(3)	16(8)	-44(4)
Trp	-1.9(7)	-3.7(15)	-0.77(12)	0.1(12)	-17.6(8)	-1.7(7)	-6(2)	-0.71(11)	0.1(12)	-17.4(8)
Tyr	-8(5)	-13(3)	-10.4(17)	6(4)	-28(6)	-19(7)	-11(3)	-9.9(18)	6(5)	-39(7)
Val1	5(2)	-5(2)	-9(2)	3(3)	-14(3)	5(3)	-5(2)	-12(3)	3(3)	-16(3)
Val2	5(2)	-18(4)	-6.0(8)	13(7)	-6(3)	6(2)	-13(4)	-4.7(8)	8(5)	-8(2)
Σ	-574(90)	-195(40)	-199(23)	301(103)	-764(50)	-574 (60)	-195(40)	-195(22)	296(105)	-765(45)

Table S7.
Full Table of averaged interaction energies in COX-I with ESD from averaging over trajectories in kJ/mol. Calculated using *tonto*, backend of *CrystalExplorer*.[32]

	Ibuprofen (**1**)					Sila-ibuprofen (**2**)				
Residue	Ele	Disp	Pol	Rep	Total	Ele	Disp	Pol	Rep	Total
Arg	-449(25)	-19.7(16)	-98(9)	134(39)	-494(12)	-465(24)	-19.1(11)	-111(8)	161(39)	-492(10)
GlyAla	-29(4)	-34(4)	-12.0(10)	31(10)	-51(7)	-28(5)	-40(5)	-14.9(15)	30(10)	-57(4)
LeuSer	-16(4)	-35(7)	-7.4(9)	25(11)	-16(5)	-18(4)	-29(7)	-8.6(12)	20(10)	-38(5)
LeuTyr	-1.7(14)	-10(3)	-1.9(4)	4(3)	-31(5)	-0.6(11)	-4.6(12)	-1.34(16)	0.4(5)	-5.4(16)
Leu	-5.4(10)	-4.5(19)	-3.2(5)	2(2)	-7.1(9)	-5.3(9)	-3.3(15)	-3.1(5)	0.7(14)	-10.2(17)
MetVal	-2(5)	-34(6)	-21(3)	27(10)	-9(5)	5(4)	-31(6)	-13(2)	22(9)	-18(5)
Phe	0.1(14)	-9(2)	-1.09(12)	8(5)	-15(4)	-0.1(14)	-10(2)	-1.13(13)	10(5)	-3.3(9)
SerLeu	-24(3)	-23(4)	-6.9(9)	19(8)	-25(4)	-26(3)	-23(4)	-10.3(13)	19(8)	-43(4)
Trp	-1.8(8)	-7(2)	-0.57(7)	0.5(26)	-17(2)	-1.7(8)	-7(2)	-0.74(10)	1(3)	-8(2)
Tyr	-103(15)	-18(2)	-45(4)	98(29)	-88(5)	-99(13)	-13.2(17)	-43(4)	95(27)	-89(5)
Val1	3(3)	-7.3(19)	-12(2)	6(4)	-14(4)	10(4)	-8(2)	-12(2)	7(4)	-1(4)
Val2	6.3(20)	-14(3)	-4.9(7)	8(5)	4(2)	2(3)	-20(4)	-6.5(8)	18(8)	-9(3)
Σ	-622(65)	-216(40)	-213(23)	364(128)	-762(56)	-626(65)	-207(38)	-225(22)	383(125)	-773(47)

6.) Free energy perturbation (FEP) calculations

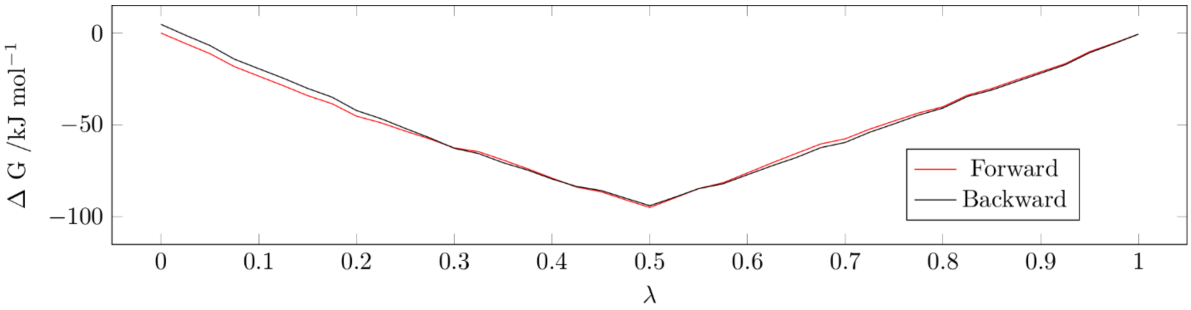


Fig. S26. Plot of ΔG against λ for COX-II, corresponding to the free energy change at each level of perturbation during the simulation.

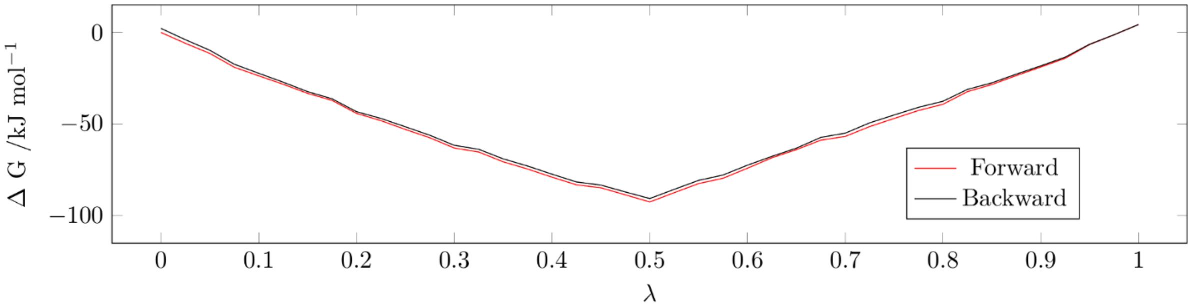


Fig. S27. Plot of ΔG against λ for COX-I, corresponding to the free energy change at each level of perturbation during the simulation. Color code as above.

7.) Toxicological investigations

Cell cultures and experimental incubations

The C6 glioma cell line was purchased from the European Collection of Authenticated Cell Cultures (Lot number: 17A034, passage number: +2). The cells were cultured as recently described in detail.[39] Briefly, the cells were cultured in cell culture medium (90% Dulbecco's modified Eagle's medium (DMEM) containing 25 *mM* glucose, with 10% fetal calf serum (FCS), 44.6 *mM* sodium bicarbonate, 1 *mM* pyruvate, 18 *units/mL* penicillin G and 18 *µg/mL* streptomycin) in 175 cm^2 flasks at 37°*C* with 10% CO_2 in humidified atmosphere in incubators from Sanyo (Osaka, Japan). Cells were subcultured after reaching approximately 80% confluency by washing the cells with 10 *mL* of pre-warmed (37°*C*) phosphate-buffered saline (PBS; 10 *mM* potassium phosphate buffer pH 7.4, 150 *mM* NaCl). Subsequently, cells were detached by incubation in 10 *mL* PBS containing 0.05 % trypsin for 5 *min* at 37°*C* with 10% CO_2. After addition of 10 *mL* cell culture medium the cell suspension was centrifuged for 5 *min* at 400 *g*, the supernatant was aspirated, and the cell pellet was resuspended in 10 *mL* fresh cell culture medium. The cell number in the suspension was determined using a Neubauer counting chamber. For experiments cells were seeded in 1 *mL* cell culture medium into wells of 24-well plates at a density of 50,000 viable cells/well and were used for experiments 24 *h* after seeding.

Cells were then exposed to ibuprofen (**1**) or sila-ibuprofen (**2**) by adding 10 *µL* of a 100-times concentrated stock in ethanol to the cell culture medium to yield final concentrations of up to 1000 *µM* of the substances and 1% ethanol. Cells were subsequently incubated for 24, 48 or 72 *h* at 37°*C* with 10% CO_2. After incubation the cell morphology was examined by phase contrast microscopy and the incubation media were harvested to determine the extracellular lactate dehydrogenase (LDH) activity as indicator for potential membrane impairment. The cells were washed and treated as described below to determine cellular LDH activity, cellular WST-1 reduction capacity and cellular protein content.

Measurement of LDH activity and protein content

The extra- and intracellular LDH activity was determined as described previously in detail.[40] Briefly, 10 *µL* of the harvested incubation medium were used for the determination of the extracellular LDH activity. For the determination of the cellular LDH activity the cells were washed twice with 1 *mL* ice-cold (4°*C*) PBS and subsequently lysed with 1% (w/v) Triton X-100 in DMEM for at least 30 *min* at *4°C*. 10 *µL* of the lysates were used to determine the cellular LDH activity.

For determination of the cellular protein content, the cells were washed twice with 1 mL ice-cold (4°*C*) PBS and the dry cells were stored at -20°C before the protein content of the cells was determined using the Lowry method[40] with bovine serum albumin as a standard protein.

Determination of extracellular lactate and cell-dependent WST1 reduction

The lactate concentration in media samples was determined as described previously.[41] Briefly, media samples harvested after a total incubation time of 72 *h* were diluted 1:5 in pure water. 10 *µL* of diluted media samples were mixed with 170 *µL* pure water in wells of a 96-well plate. Subsequently, 180 *µL* of freshly prepared reaction mixture (5.6 *mM* NAD+, 37.7 *units/mL* LDH, 3.89 *units/mL* glutamate pyruvate transaminase (GPT) in 500 *mM* glutamate/KOH buffer, pH 8.9) were added to each well and the plate was incubated for 90 *min* at 37°*C* in the humidified atmosphere of an incubator and the absorbance at 340 *nm* was measured in a Sunrise microtiter plate spectrophotometer (Tecan, Crailsheim, Germany).

The WST-1 reduction of the C6 glioma cells was determined using a modification of a method described recently in detail.[42] After the incubation the incubation medium was aspirated and the cells were washed twice with 0.5 *mL* pre-warmed (37°C) glucose-free incubation buffer (IB: 20 *mM* HEPES, 145 *mM* NaCl, 5.4 *mM* KCl, 1.8 *mM* $CaCl_2$, 1 *mM* $MgCl_2$, 0.8 *mM* Na_2HPO_4, pH 7.4) and subsequently incubated with 200 *µL* IB containing 5 *mM* glucose, 400 *µM* WST-1 and 50 *µM* menadione. After 30 *min* incubation at 37°*C*, 50 *µL* of the incubation medium was taken, mixed with 150 *µL* pure water in wells of a microtiter plate and the absorbance at 405 *nm* was measured in a Sunrise microtiter plate spectrophotometer (Tecan, Crailsheim, Germany).

Materials

Dulbecco's modified Eagle's medium (DMEM) and penicillin/streptomycin solution were purchased from Invitrogen-Gibco (Darmstadt, Germany). Fetal calf serum (FCS), menadione, ethanol and Triton X-100 were obtained from Sigma (Steinheim, Germany). DMSO was obtained from VWR Chemicals (Darmstadt, Germany). Trypsin solution was purchased from Biochrom (Berlin, Germany). WST-1 was obtained from Dojindo (Munich, Germany). 2-[(4-bromomethyl)phenyl]propionic acid and dimethylchlorosilane were obtained from abcr (Karlsruhe, Germany) and ibuprofen (**1**) sodium salt was purchased from Merck (Darmstadt, Germany). The enzymes lactate dehydrogenase (LDH) and glutamate pyruvate transaminase (GPT) were obtained from Roche Diagnostics (Mannheim, Germany)

Other chemicals of the highest purity available were purchased from Merck (Darmstadt, Germany), Applichem (Darmstadt, Germany), Fluka (Buchs, Switzerland), Roth (Karlsruhe, Germany) or Riedel-de-Haën (Seelze, Germany). Sterile 175 cm^2 flasks were obtained from VWR (Darmstadt, Germany). Sterile 24-well cell culture plates and unsterile 96-well microtiter plates were obtained from Sarstedt (Nümbrecht, Germany).

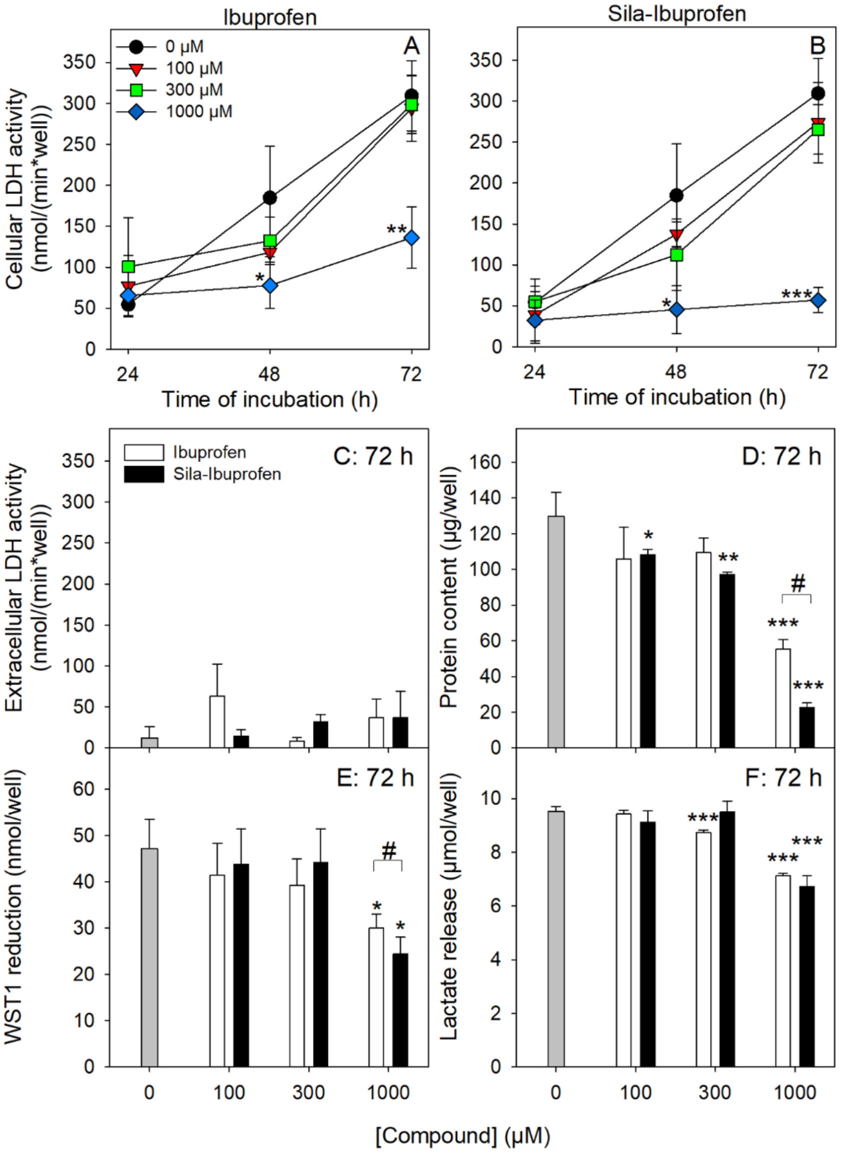


Fig. S28.
Time- and concentration-dependent effects of ibuprofen (**1**) or sila-ibuprofen (**2**) on the viability and metabolic activity of C6 glioma cells. The cells were incubated for up to 72 *h* (A, B) or for 72 *h* (C-F) with **1** or **2** in the concentrations indicated before the cellular (A, B) and extracellular (C) LDH activity, the cellular protein content (D), the cell-dependent WST1 reduction (E) and the lactate release (F) were determined. The data represent means ± SD of values obtained in 3 experiments performed on different passages of C6 cells. Significant differences compared to the control condition (0 *µM* of the compound) were calculated by ANOVA (followed by the Bonferroni post hoc test) and are indicated by *p<0.05, **p<0.01 and ***p<0.001. Significant differences between the data obtained for incubations with **1** or **2** in a given concentration were calculated by the paired t-test and are indicated by #p<0.05.

8. Enzyme activity measurements to determine IC_{50} values

Table S8.
Inhibition (% of control) of COX-I and COX-II by **1** and **2**, one repetition each

	Ibuprofen (**1**)				Sila-ibuprofen (**2**)			
Concentration /µM	COX-I		COX-II		COX-I		COX-II	
0.1	16.9	11.0	29.3	28.9	19.1	23.1	16.7	16.0
0.316	4.8	7.0	39.8	44.3	26.3	26.5	18.3	23.8
1.0	-0.6	22.9	54.8	53.3	19.0	20.8	34.6	28.3
3.16	5.0	11.1	54.0	57.9	19.5	24.7	47.5	45.8
10	6.4	20.1	68.1	67.6	33.9	35.7	57.2	53.3
31.6	54.7	72.6	85.5	79.6	57.4	64.9	70.6	74.9
100	94.5	88.7	90.6	94.4	78.0	86.0	99.1	91.4

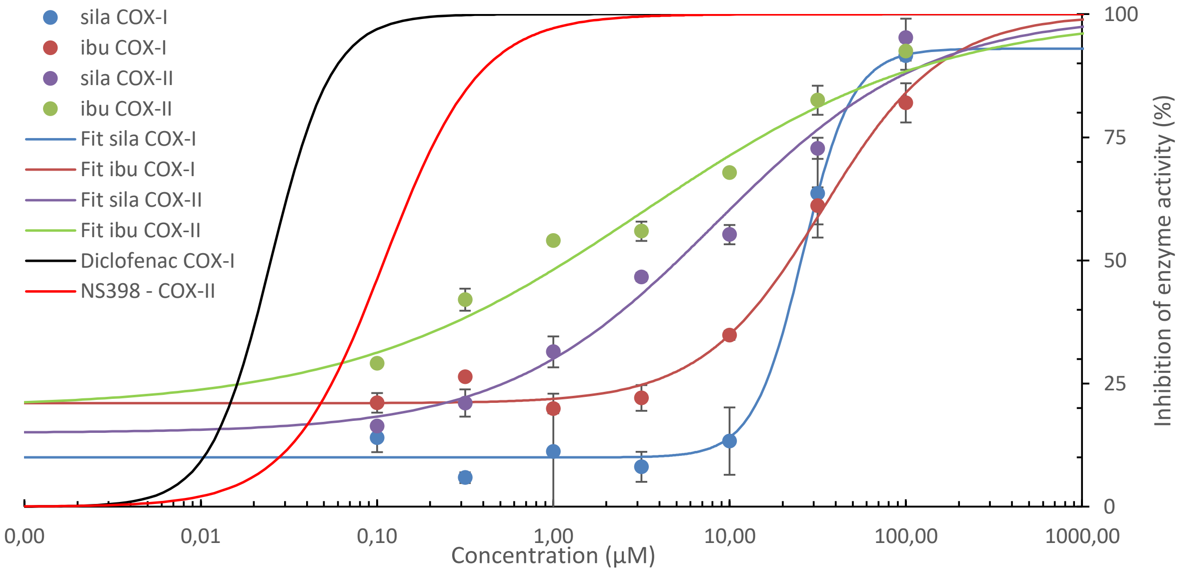


Fig. S29.
Concentration-dependent inhibition of COX-I and COX-II by ibuprofen (**1**, "ibu") and sila-ibuprofen (**2**, "sila"). Shown are also the fitted regression curves for the compounds investigated as well as for the reference substances diclofenac and NS398 used for validation of the test systems.

9. References to methods and software used in the bonding analysis

- Quantum Theory of Atoms in Molecules (QTAIM)[43]: software AIMAll.[44]
- Electron Localizability Indicator (ELI)[45]; Raub-Jansen index[46]: software DGrid-5.0.[47]
- Natural bond orbitals (NBOs) including natural population analysis (NPA) and natural localized molecular orbitals (NLMOs)[48]; natural resonance theory (NRT)[49]: software NBO-7.0.[50]
- Electrostatic potential (ESP) plotted using VMD[51] based on a grid file calculated with cuQCT (home-written software, unpublished).

10. Captions for other Supporting Information

Movie S1.
Guided visual representation of sila-ibuprofen – COX-II complex (ball and stick representation for sila-ibuprofen and NewCartoon[52] for COX-II 0:00 – 0:10), amino acid residues of importance (licorice representation, 0:10 – 0:33) and aNCI isosurfaces (starting from 0:19) with color code; then a side by side comparison of sila-ibuprofen (**2**,left) and ibuprofen (**1**,right) aNCI plots. Representations were created using VMD 1.9.3[51] and the video rendered using Blender 2.79b.[53] Geometries used for the visualization of the atom positions correspond to the last frame of a 1 ns production run, the aNCI is averaged over 1000 frames of this run.

Data S1. (separate file: interaction_energies.xlsx)

Interaction energies for each frame and the analysis of average and wRMSD, separated by contribution (Ele, Disp, Pol and Rep) and in total for sila-ibuprofen and ibuprofen, respectively, as well as a tab with summary for both.

Data S2. (separate file: bond_length_plots.xlsx)

O-N bond lengths in each frame of MD for sila-ibuprofen and ibuprofen, respectively, both subunits.

Data S3. (separate files: bond_order.xlsx)

Bond properties from QTAIM and NBO analysis for all bonds and ELI-D for selected bonds.

Data S4. (separate files: *.cif, *.fcf and *.pdf)
Crystallographic information files (CIFs) of **1** and **2**, including measured reflection intensities and checkcif reports.

Data S5. (separate file: biochemical_data.csv)
Names, SMILES notation and IC_{50} values of ibuprofen **1** and sila-ibuprofen **2**.

11. References for Supporting Information

(38) Mayne, C. G.; Saam, J.; Schulten, K.; Tajkhorshid, E.; Gumbart, J. C. Rapid parameterization of small molecules using the force field toolkit. *J. Comput. Chem.* **2013**, *34*, 2757-2770.
(39) Joshi, A.; Rastedt, W.; Faber, K.; Schultz, A. G.; Bulcke, F.; Dringen, R. Uptake and toxicity of copper oxide nanoparticles in C6 glioma cells. *Neurochem. Res.* **2016**, *41*, 3004.
(40) Lowry, O. H.; Rosebrough, N. J.; Farr, A. L.; Randall, R. J. Protein measurement with the Folin phenol reagent, *J. Biol. Chem.* **1951**, *193*, 265-275.
(41) Tulpule, K.; Hohnholt, M. C.; Hirrlinger, J.; Dringen, R. “Primary cultures of astrocytes and neurons as model systems to study the metabolism and metabolite export from brain cells” in *Brain Energy Metabolism*. Springer, New York **2014,** 45-72.
(42) Stapelfeldt, K.; Ehrke, E.; Steinmeier, J.; Rastedt, W.; Dringen, R. Menadione-mediated WST1 reduction assay for the determination of metabolic activity of cultured neural cells, *Anal. Biochem.* **2017**, 538, 42-52.
(43) Bader, R. F. W. *Atoms in Molecules: A Quantum Theory*, Clarendon Press, Oxford, **1994**.
(44) Keith, T. A., *AIMAll* (Version 19.10.12), TK Gristmill Software, Overland Park KS, USA, **2019** (aim.tkgristmill.com).
(45) Kohout, M. Bonding indicators from electron pair density functionals. *Faraday Discuss* **2007**, *135*, 43-54.
(46) Raub, S.; Jansen, G. A quantitative measure of bond polarity from the electron localization function and the theory of atoms in molecules. *Theor. Chem. Acc.* **2001**, *106*, 223 – 232.
(47) Kohout, M. *DGrid*, Radebeul, Germany **2011**.
(48) Weinhold, F.; Landis, C. R.**;** Glendening, E. D. What is NBO analysis and how is it useful? *Int. Rev. Phys. Chem.* **2016**, *35*, 399-440.
(49) Glendening, E.D.; Landis, C. R.; Weinhold, F. Resonance theory reboot. *J. Am. Chem. Soc.* **2019**, *141*, 4156-4166.
(50) Glendening, E.D.; Badenhoop, J, K.; Reed, A. E.; Carpenter, J. E.; Bohmann, J. A.; Morales, C. M.; Karafiloglou, P.; Landis, C. R.; Weinhold, F. *NBO 7.0*, Theoretical Chemistry Institute, University of Wisconsin, Madison **2018**.
(51) Humphrey, W.; Dalke, A.; Schulten, K. VMD – Visual molecular dynamics. *J. Mol. Graph.* **1996**, *14*, 33-38.
(52) Frishman, D.; Argos, P. Knowledge-based secondary structure assignment. *Proteins: structure, function and genetics* **1995**, *23*, 566-579.
(53) Blender Online Community, Blender – a 3D modelling and rendering package (blender.org).

Bibliography

[1] S. Galli, *J. Chem. Educ.* **2014**, *91*(12), 2009–2012.

[2] J. D. Watson, F. H. C. Crick, *Nature* **1953**, *171*(4356), 737–738.

[3] H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov, P. E. Bourne, *Nucleic Acids Res.* **2000**, *28*(1), 235–242.

[4] K. S. Novoselov, A. K. Geim, S. V. Morozov, D. Jiang, Y. Zhang, S. V. Dubonos, I. V. Grigorieva, A. A. Firsov, *Science* **2004**, *306*(5696), 666–669.

[5] D. Shechtman, C. I. Lang, *MRS Bull.* **1997**, *22*(11), 40–42.

[6] IUCr Executive Committee, *Acta Cryst.* **1992**, *A48*(6), 928.

[7] M. v. Laue, *Die Interferenzen von Röntgen-und Elektronenstrahlen: fünf Vorträge*, Springer, **1935**.

[8] W. L. Bragg, *Proc. R. Soc. Lond.* **1913**, *A 89*, 248–277.

[9] W. H. Bragg, W. L. Bragg, *Proc. R. Soc. Lond.* **1913**, *A 88*, 428–438.

[10] W. L. Bragg, *Proc. R. Soc. Lond.* **1914**, *A 89*, 468–489.

[11] P. P. Ewald, *Z. Krist.-Cryst. Mater.* **1921**, *56*(1), 129–156.

[12] P. Debye, P. Scherrer, *Nachr. Ges. Wiss. Göttingen* **1916**, *29*, 1–15.

[13] G. W. Stewart, *Phys. Rev.* **1928**, *32*(4), 558–563.

[14] I. Fankuchen, M. L. Anson, J. T. Edsall, in *Advances in Protein Chemistry*, vol. 2, Academic Press, **1945**, 387–405.

[15] B. E. Warren, *J. Appl. Phys.* **1941**, *12*(5), 375–384.

[16] W. B. Yelon, F. K. Ross, *Nucl. Instrum. Methods Phys. Res.* **1982**, *193*(1), 285–292.

[17] S. W. Peterson, H. A. Levy, *Acta Cryst.* **1957**, *10*(1), 70–76.

[18] T. Gruene, J. T. C. Wennmacher, C. Zaubitzer, J. J. Holstein, J. Heidler, A. Fecteau-Lefebvre, S. DeCarlo, E. Müller, K. N. Goldie, I. Regeni, T. Li, G. Santiso-Quinones, G. Steinfeld, S. Handschin, E. vanGenderen, J. A. vanBokhoven, G. H. Clever, R. Pantelic, *Angew. Chem. Int. Edit.* **2018**, *57*(50), 16313–16317.

[19] W. Cochrane, *P. Phys. Soc.* **1936**, *48*(5), 723–735.

[20] J. D. Dunnitz, *X-Ray Analysis and the Structure of organic Molecules*, Verlag Helvetica Chimica Acta, VCH, **1995**.

[21] W. Massa, *Kristallstrukturbestimmung*, vol. 7, Springer, Wiesbaden, **2011**.

[22] C. Brönnimann, P. Trüb, in *Synchrotron Light Sources and Free-Electron Lasers: Accelerator Physics, Instrumentation and Science Applications*, (Published by E. J. Jaeschke, S. Khan, J. R. Schneider, J. B. Hastings), Springer International Publishing, Cham, **2016**, 995–1027.

[23] T. Hirono, H. Toyokawa, Y. Furukawa, M. Kawase, T. Ohata, S. Wu, H. Ikeda, G. Sato, T. Takahashi, S. Watanabe, *Nucl. Instrum. Methods Phys. Res., Sect. A* **2013**, *731*, 64–67.

[24] J. D. Dunitz, P. Seiler, *Acta Cryst.* **1973**, *B29*(3), 589–595.

[25] E. T. Whittaker, G. Robinson, *Calculus of observations: Treatise on Numerical Mathematics 3rd Edition.*, Blackie and Sons Limited, London, **1940**.

[26] A. M. Legendre, *Nouvelles méthodes pour la determination des orbites des cometes: avec un suplement contenant divers erfectionnemens de ces methodes et leur application aux deux cometes de 1805*, Courcier, **1806**.

[27] A. J. C. Wilson, *Acta Cryst.* **1976**, *A32*(6), 994–996.

[28] D. C. Phillips, D. Rogers, A. J. C. Wilson, *Acta Cryst.* **1950**, *3*(5), 398–399.

[29] A. J. C. Wilson, *Acta Cryst.* **1950**, *3*(4), 258–261.

[30] A. J. C. Wilson, *Acta Cryst.* **1950**, *3*(5), 397–398.

[31] O. V. Dolomanov, L. J. Bourhis, R. J. Gildea, J. A. K. Howard, H. Puschmann, *J. Appl. Crystallogr.* **2009**, *42*(2), 339–341.

[32] A. L. Spek, *Acta Cryst.* **2015**, *C71*(1), 9–18.

[33] P. van der Sluis, A. L. Spek, *Acta Cryst.* **1990**, *A46*(3), 194–201.

[34] P. Debye, *Phys. Z.* **1930**, *31*, 419.

[35] R. Stewart, *Acta Cryst.* **1975**, *A31*(s3), s218–s218.

[36] J. M. Robertson, W. H. Bragg, *P. Phys. Soc. Lond. A* **1934**, *146*(857), 473–482.

[37] S. W. H. Bragg, *P. Phys. Soc. Lond.* **1921**, *34*(1), 98–103.

[38] W. Cochran, *Acta Cryst.* **1948**, *1*(3), 138–142.

[39] E. W. Hughes, *J. Am. Chem. Soc.* **1941**, *63*(6), 1737–1752.

[40] A. H. Compton, *Nature* **1915**, *95*(2378), 343–344.

[41] D. T. Cromer, J. B. Mann, *Acta Cryst.* **1968**, *A24*(2), 321–324.

[42] E. Rutherford, *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* **1911**, *21*(125), 669–688.

[43] N. Bohr, *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* **1913**, *26*(153), 1–25.

[44] N. Bohr, *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* **1913**, *26*(153), 476–502.

[45] N. Bohr, *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* **1913**, *26*(155), 857–875.

[46] H. Svendsen, J. Overgaard, R. Busselez, B. Arnaud, P. Rabiller, A. Kurita, E. Nishibori, M. Sakata, M. Takata, B. B. Iversen, *Acta Cryst.* **2010**, *A66*(4), 458–469.

[47] A. Fischer, D. Tiana, W. Scherer, K. Batke, G. Eickerling, H. Svendsen, N. Bindzus, B. B. Iversen, *J. Phys. Chem. A* **2011**, *115*(45), 13061–13071.

[48] A. M. O'Connell, A. I. M. Rae, E. N. Maslen, *Acta Cryst.* **1966**, *21*(2), 208–219.

[49] F. L. Hirshfeld, *Acta Cryst.* **1971**, *B27*(4), 769–781.

[50] L. J. Bourhis, O. V. Dolomanov, R. J. Gildea, J. A. K. Howard, H. Puschmann, *Acta Cryst.* **2015**, *A71*(1), 59–75.

[51] D. Zankov, T. I. Madzhidov, A. Rakhimbekova, T. R. Gimadiev, R. I. Nugmanov, M. A. Kazymova, I. I. Baskin, A. Varnek, *J. Chem. Inf. Model.* **2019**, *59*(11), 4569–4576.

[52] D. T. Stanton, *J. Chem. Inf. Comput. Sci.* **2000**, *40*(1), 81–90.

[53] O. Mekenyan, D. Bonchev, N. Trinajstić, *Int. J. Quant. Chem.* **1980**, *18*(2), 369–380.

[54] F. Neese, *WIREs Comput. Mol. Sci.* **2012**, *2*(1), 73–78.

[55] F. Neese, *WIREs Comput. Mol. Sci.* **2018**, *8*(1), e1327.

[56] W. Humphrey, A. Dalke, K. Schulten, et al., *J. Mol. Graph.* **1996**, *14*(1), 33–38.

[57] D. Watkin, *Acta Cryst.* **2000**, *B56*(5), 747–749.

[58] D. Watkin, *J. Appl. Crystallogr.* **2008**, *41*(3), 491–522.

[59] R. F. Stewart, E. R. Davidson, W. T. Simpson, *J. Chem. Phys.* **1965**, *42*(9), 3175–3187.

[60] J. C. Hanson, L. C. Sieker, L. H. Jensen, *Acta Cryst.* **1973**, *B29*(4), 797–808.

[61] R. Taylor, O. Kennard, *Acc. Chem. Res.* **1984**, *17*(9), 320–326.

[62] R. Taylor, O. Kennard, W. Versichel, *J. Am. Chem. Soc.* **1983**, *105*(18), 5761–5766.

[63] F. H. Allen, *Acta Cryst.* **1986**, *B42*(5), 515–522.

[64] F. H. Allen, I. J. Bruno, *Acta Cryst.* **2010**, *B66*(3), 380–386.

[65] L. A. Malaspina, A. J. Edwards, M. Woiska, D. Jayatilaka, M. J. Turner, J. R. Price, R. Herbst-Irmer, K. Sugimoto, E. Nishibori, S. Grabowsky, *Cryst. Growth Des.* **2017**, *17*(7), 3812–3825.

[66] N. K. Hansen, P. Coppens, *Acta Cryst.* **1978**, *A34*(6), 909–921.

[67] F. L. Hirshfeld, *Isr. J. Chem.* **1964**, *2*(3), 87–90.

[68] F. L. Hirshfeld, *Acta Cryst.* **1971**, *B27*(4), 769–781.

[69] F. L. Hirshfeld, *Isr. J. Chem.* **1977**, *16*(23), 226–229.

[70] R. Chen, P. Trucano, R. F. Stewart, *Acta Cryst.* **1977**, *A33*(5), 823–828.

[71] K. Kurki-Suonio, *Acta Cryst.* **1968**, *A24*(3), 379–390.

[72] K. Kurki-Suonio, *Isr. J. Chem.* **1977**, *16*(23), 132–136.

[73] P. Coppens, *Angew. Chem. Int. Edit.* **2005**, *44*(42), 6810–6811.

[74] T. S. Koritsanszky, P. Coppens, *Chem. Rev.* **2001**, *101*(6), 1583–1628.

[75] U. Flierler, D. Stalke, *Electron Density and Chemical Bonding I*, Springer-Verlag GmbH, Berlin, Heidelberg, **2012**, 1–20.

[76] D. Stalke, *Electron Density and Chemical Bonding II: Theoretical Charge Density Studies*, vol. 147, Springer, **2012**.

[77] D. Stalke, *Chem.–Eur. J.* **2011**, *17*(34), 9264–9278.

[78] B. Dittrich, C. B. Hübschle, K. Pröpper, F. Dietrich, T. Stolper, J. J. Holstein, *Acta Cryst.* **2013**, *B69*(2), 91–104.

[79] S. Domagała, B. Fournier, D. Liebschner, B. Guillot, C. Jelsch, *Acta Cryst.* **2012**, *A68*(3), 337–351.

[80] K. N. Jarzembska, P. M. Dominiak, *Acta Cryst.* **2012**, *A68*(1), 139–147.

[81] M. L. Chodkiewicz, S. Migacz, W. Rudnicki, A. Makal, J. A. Kalinowski, N. W. Moriarty, R. W. Grosse-Kunstleve, P. V. Afonine, P. D. Adams, P. M. Dominiak, *J. Appl. Crystallogr.* **2018**, *51*(1), 193–199.

[82] R. Herbst-Irmer, J. Henn, J. J. Holstein, C. B. Hübschle, B. Dittrich, D. Stern, D. Kratzert, D. Stalke, *J. Phys. Chem. A* **2013**, *117*(3), 633–641.

[83] K. Meindl, R. Herbst-Irmer, J. Henn, *Acta Cryst.* **2010**, *A66*(3), 362–371.

[84] P. R. Mallinson, T. Koritsanszky, E. Elkaim, N. Li, P. Coppens, *Acta Cryst.* **1988**, *A44*(3), 336–343.

[85] A. Poulain, E. Wenger, P. Durand, K. N. Jarzembska, R. Kamiński, P. Fertey, M. Kubicki, C. Lecomte, *IUCrJ* **2014**, *1*(2), 110–118.

[86] C. K. Johnson, H. A. Levy, *International Tables for X-ray Crystallography, vol. IV* **1974**, 314–319.

[87] A. Volkov, P. Macchi, L. J. Farrugia, C. Gatti, P. Mallinson, T. Richter, T. Koritsanszky, **2016**.

[88] C. Jelsch, B. Guillot, C. Iordache, **2019**.

[89] B. Engels, T. C. Schmidt, C. Gatti, T. Schirmeister, R. F. Fink, in *Electron Density and Chemical Bonding II*, Springer, **2011**, 47–97.

[90] P. Coppens, Y. Abramov, M. Carducci, B. Korjov, I. Novozhilova, C. Alhambra, M. R. Pressprich, *J. Am. Chem. Soc.* **1999**, *121*(11), 2585–2593.

[91] J. M. Bak, P. M. Dominiak, C. C. Wilson, K. Woźniak, *Acta Cryst.* **2009**, *A65*(6), 490–500.

[92] M. Woińska, D. Jayatilaka, B. Dittrich, R. Flaig, P. Luger, K. Woźniak, P. M. Dominiak, S. Grabowsky, *ChemPhysChem* **2017**, *18*(23), 3334–3351.

[93] K. K. Jha, B. Gruza, P. Kumar, M. L. Chodkiewicz, P. M. Dominiak, *Acta Cryst.* **2020**, *B76*(3), 296–306.

[94] E. T. Jaynes, *Phys. Rev.* **1957**, *106*, 620–630.

[95] E. T. Jaynes, *Phys. Rev.* **1957**, *108*, 171–190.

[96] A. Ben-Naim, *Introduction*, World Scientific Publishing Co. Pte. Ltd., chap. 1.3, **2008**, 19–31.

[97] S. van Smaalen, L. Palatinus, M. Schneider, *Acta Cryst.* **2003**, *A59*(5), 459–469.

[98] H. Tanaka, M. Takata, E. Nishibori, K. Kato, T. Iishi, M. Sakata, *J. Appl. Crystallogr.* **2002**, *35*(2), 282–286.

[99] E. Nishibori, T. Nakamura, M. Arimoto, S. Aoyagi, H. Ago, M. Miyano, T. Ebisuzaki, M. Sakata, *Acta Cryst.* **2008**, *D64*(3), 237–247.

[100] P. Hohenberg, W. Kohn, *Phys. Rev.* **1964**, *136*, B864–B871.

[101] D. Bagayoko, *AIP Adv.* **2014**, *4*(12), 127104.

[102] W. L. Clinton, J. Nakhleh, F. Wunderlich, *Phys. Rev.* **1969**, *177*, 1–6.

[103] W. L. Clinton, A. J. Galli, L. J. Massa, *Phys. Rev.* **1969**, *177*, 7–13.

[104] W. L. Clinton, G. A. Henderson, J. V. Prestia, *Phys. Rev.* **1969**, *177*, 13–18.

[105] W. L. Clinton, G. B. Lamers, *Phys. Rev.* **1969**, *177*, 19–27.

[106] W. L. Clinton, A. J. Galli, G. A. Henderson, G. B. Lamers, L. J. Massa, J. Zarur, *Phys. Rev.* **1969**, *177*, 27–33.

[107] W. L. Clinton, L. J. Massa, *Int. J. Quant. Chem.* **1972**, *6*(3), 519–523.

[108] W. L. Clinton, L. J. Massa, *Phys. Rev. Lett.* **1972**, *29*, 1363–1366.

[109] W. L. Clinton, C. A. Frishberg, L. J. Massa, P. A. Oldfield, *Int. J. Quant. Chem.* **1973**, *7*(S7), 505–514.

[110] L. Massa, M. Goldberg, C. Frishberg, R. F. Boehme, S. J. L. Placa, *Phys. Rev. Lett.* **1985**, *55*, 622–625.

[111] L. Massa, L. Huang, J. Karle, *Int. J. Quant. Chem.* **1995**, *56*(S29), 371–384.

[112] L. Huang, L. Massa, J. Karle, *International Journal of Quantum Chemistry* **2005**, *103*(6), 808–817.

[113] L. Huang, L. Massa, J. Karle, *P. Natl. Acad. Sci. USA* **2005**, *102*(36), 12690–12693.

[114] L. Huang, L. Massa, J. Karle, *P. Natl. Acad. Sci. USA* **2006**, *103*(5), 1233–1237.

[115] L. Huang, L. Massa, J. Karle, *J. Mol. Struct.* **1999**, *474*(1), 9–12.

[116] L. Massa, C. F. Matta, *J. Comput. Chem.* **2018**, *39*(17), 1021–1028.

[117] S. Grabowsky, A. Genoni, H.-B. Bürgi, *Chem. Sci.* **2017**, *8*(6), 4159–4176.

[118] A. Genoni, L. Buinský, N. Claiser, J. Contreras-García, B. Dittrich, P. M. Dominiak, E. Espinosa, C. Gatti, P. Giannozzi, J.-M. Gillet, D. Jayatilaka, P. Macchi, A. . Madsen, L. Massa, C. F. Matta, K. M. Merz, P. N. H. Nakashima, H. Ott, U. Ryde, K. Schwarz, M. Sierka, S. Grabowsky, *Chem.–Eur. J.* **2018**, *24*(43), 10881–10905.

[119] C. F. Matta, L. Massa, *Future Med. Chem.* **2018**, *10*(13), 1525–1527.

[120] L. Massa, C. F. Matta, *Can. J. Chem.* **2018**, *96*(7), 599–605.

[121] W. Polkosnik, C. F. Matta, L. Huang, L. Massa, *Int. J. Quant. Chem.* **2019**, *119*(24), e25986.

[122] A. Genoni, P. Macchi, *Crystals* **2020**, *10*(6), 473.

[123] A. A. Korlyukov, Y. V. Nelyubina, *Russ. Chem. Rev.* **2019**, *88*(7), 677–716.

[124] P. Macchi, P. Coppens, *Acta Cryst.* **2001**, *A57*(6), 656–662.

[125] S. C. Capelli, H.-B. Bürgi, B. Dittrich, S. Grabowsky, D. Jayatilaka, *IUCrJ* **2014**, *1*(5), 361–379.

[126] D. Jayatilaka, B. Dittrich, *Acta Cryst.* **2008**, *A64*(3), 383–393.

[127] D. Jayatilaka, D. J. Grimwood, *Acta Cryst.* **2001**, *A57*(1), 76–86.

[128] D. J. Grimwood, D. Jayatilaka, *Acta Cryst.* **2001**, *A57*(1), 87–100.

[129] D. Jayatilaka, P. Munshi, M. J. Turner, J. A. K. Howard, M. A. Spackman, *Phys. Chem. Chem. Phys.* **2009**, *11*, 7209–7218.

[130] S. Grabowsky, P. Luger, J. Buschmann, T. Schneider, T. Schirmeister, A. N. Sobolev, D. Jayatilaka, *Angew. Chem. Int. Edit.* **2012**, *51*(27), 6776–6779.

[131] F. L. Hirshfeld, *Theor. Chim. Acta* **1977**, *44*(2), 129–138.

[132] B. Meyer, A. Genoni, *J. Phys. Chem. A* **2018**, *122*(45), 8965–8981.

[133] M. Sironi, A. Genoni, M. Civera, S. Pieraccini, M. Ghitti, *Theor. Chem. Acc.* **2007**, *117*(5), 685–698.

[134] L. Malaspina, E. K. Wieduwilt, J. Bergmann, F. Kleemiss, B. Meyer, M. F. Ruiz-López, R. Pal, E. Hupf, J. Beckmann, R. O. Piltz, A. J. Edwards, S. Grabowsky, A. Genoni, *J. Phys. Chem. Lett.* **2019**, *10*(22), 6973–6982.

[135] S. Grabowsky, A. Genoni, S. P. Thomas, D. Jayatilaka, *The Advent of Quantum Crystallography: Form and Structure Factors from Quantum Mechanics for Advanced Structure Refinement and Wavefunction Fitting*, Springer Nature Switzerland AG, **2020**.

[136] M. E. Wall, *IUCrJ* **2016**, *3*(4), 237–246.

[137] L. Huang, C. Matta, L. Massa, *Struct. Chem.* **2015**, *26*(5), 1433–1442.

[138] L. Huang, L. Massa, J. Karle, Quantum kernels and quantum crystallography: Applications in biochemistry, **2010**.

[139] P. Macchi, J.-M. Gillet, F. Taulelle, J. Campo, N. Claiser, C. Lecomte, *IUCrJ* **2015**, *2*(4), 441–451.

[140] M. Woińska, S. Grabowsky, P. M. Dominiak, K. Woźniak, D. Jayatilaka, *Sci. Adv.* **2016**, *2*(5).

[141] L. Pauling, *J. Am. Chem. Soc.* **1932**, *54*(9), 3570–3582.

[142] W. Kohn, *Phys. Rev. Lett.* **1996**, *76*, 3168–3171.

[143] W. Yang, *Phys. Rev. Lett.* **1991**, *66*(11), 1438.

[144] W. Yang, T.-S. Lee, *J. Chem. Phys.* **1995**, *103*(13), 5674–5678.

[145] G. Galli, M. Parrinello, *Phys. Rev. Lett.* **1992**, *69*(24), 3547.

[146] W. Kohn, *Chem. Phys. Lett.* **1993**, *208*(3-4), 167–172.

[147] P. Ordejón, D. A. Drabold, R. M. Martin, M. P. Grumbach, *Phys. Rev. B* **1995**, *51*(3), 1456.

[148] F. Mauri, G. Galli, R. Car, *Phys. Rev. B* **1993**, *47*(15), 9973.

[149] X.-P. Li, R. W. Nunes, D. Vanderbilt, *Phys. Rev. B* **1993**, *47*(16), 10891.

[150] C. M. Goringe, E. Hernández, M. J. Gillan, I. J. Bush, *Comput. Phys. Commun.* **1997**, *102*(1), 1–16.

[151] A. Fabrizio, A. Grisafi, B. Meyer, M. Ceriotti, C. Corminboeuf, *Chem. Sci.* **2019**, *10*(41), 9424–9432.

[152] G. Macetti, A. Genoni, *J. Phys. Chem. A* **2019**, *123*(43), 9420–9428.

[153] D. Jayatilaka, D. J. Grimwood, in *Lect. Notes Comput. Sc.*, (Published by P. M. A. Sloot, D. Abramson, A. V. Bogdanov, Y. E. Gorbachev, J. J. Dongarra, A. Y. Zomaya), Springer Berlin Heidelberg, Berlin, Heidelberg, **2003**, 142–151.

[154] P. Pollak, F. Weigend, *J. Chem. Theory Comput.* **2017**, *13*(8), 3696–3705.

[155] F. E. Jorge, A. Canal Neto, G. G. Camiletti, S. F. Machado, *J. Chem. Phys.* **2009**, *130*(6), 064108.

[156] E. V. R. de Castro, F. E. Jorge, *J. Chem. Phys.* **1998**, *108*(13), 5225–5229.

[157] J. Christensen, P. N. Horton, C. S. Bury, J. L. Dickerson, H. Taberman, E. F. Garman, S. J. Coles, *IUCrJ* **2019**, *6*(4), 703–713.

[158] S.-K. Son, H. N. Chapman, R. Santra, *Phys. Rev. Lett.* **2011**, *107*, 218102.

[159] M. R. Howells, T. Beetz, H. N. Chapman, C. Cui, J. M. Holton, C. J. Jacobsen, J. Kirz, E. Lima, S. Marchesini, H. Miao, D. Sayre, D. A. Shapiro, J. C. H. Spence, D. Starodub, *J. Electron Spectrosc.* **2009**, *170*(1), 4–12.

[160] R. Henderson, *Q. Rev. Biophys.* **1995**, *28*(2), 171–193.

[161] E. Garman, *Curr. Opin. Struct. Biol.* **2003**, *13*(5), 545–551.

[162] R. Henderson, *Proc. R. Soc. Lond. B Biol. Sci.* **1990**, *241*(1300), 6–8.

[163] R. B. G. Ravelli, S. M. McSweeney, *Structure* **2000**, *8*(3), 315–328.

[164] E. Prince, L. W. Finger, J. H. Konnert, in *International Tables for Crystallography*, International Union of Crystallography, **2006**, 694–701.

[165] G. M. Sheldrick, *Acta Cryst.* **2008**, *A64*(1), 112–122.

[166] G. M. Sheldrick, *Acta Cryst.* **2015**, *C71*(1), 3–8.

[167] F. Bloch, *Z. Phys.* **1929**, *52*(7-8), 555–600.

[168] J. C. Slater, *Phys. Rev.* **1937**, *51*, 846–851.

[169] J. Callaway, *Phys. Rev.* **1955**, *97*, 933–936.

[170] J. C. Slater, *Phys. Rev.* **1953**, *92*, 603–608.

[171] V. Heine, *P. Phys. Soc. Lond. A* **1957**, *240*(1222), 354–360.

[172] M. Fugel, M. F. Hesse, R. Pal, J. Beckmann, D. Jayatilaka, M. J. Turner, A. Karton, P. Bultinck, G. S. Chandler, S. Grabowsky, *Chem.–Eur. J.* **2018**, *24*(57), 15275–15286.

[173] M. Fugel, L. A. Malaspina, R. Pal, S. P. Thomas, M. W. Shi, M. A. Spackman, K. Sugimoto, S. Grabowsky, *Chem.–Eur. J.* **2019**, *25*(26), 6523–6532.

[174] M. Fugel, J. Beckmann, D. Jayatilaka, G. V. Gibbs, S. Grabowsky, *Chem.–Eur. J.* **2018**, *24*(23), 6248–6261.

[175] M. Fugel, M. V. Ponomarenko, M. F. Hesse, L. A. Malaspina, F. Kleemiss, K. Sugimoto, A. Genoni, G.-V. Röschenthaler, S. Grabowsky, *Dalton Trans.* **2019**, *48*, 16330–16339.

[176] S. Grabowsky, *Complementary Bonding Analysis*, De Gruyter, Berlin, Boston, **15 Mar. 2021**.

[177] R. F. W. Bader, T. T. Nguyen-Dang, *Adv. Quantum Chem.* **1981**, *14*, 63–124.

[178] R. F. W. Bader, *Atoms in Molecules: A Quantum Theory*, Clarendon Press, Oxford, **1994**.

[179] R. F. W. Bader, A. Streitwieser, A. Neuhaus, K. E. Laidig, P. Speers, *J. Am. Chem. Soc.* **1996**, *118*(21), 4959–4965.

[180] C. F. Matta, R. J. Boyd, *An Introduction to the Quantum Theory of Atoms in Molecules*, John Wiley & Sons, Ltd, chap. 1, **2007**, 1–34.

[181] C. Gatti, F. Cargnoni, L. Bertini, *J. Comput. Chem.* **2003**, *24*(4), 422–436.

[182] R. F. Bader, C. Gatti, *Chem. Phys. Lett.* **1998**, *287*(3), 233 – 238.

[183] E. R. Johnson, S. Keinan, P. Mori-Sánchez, J. Contreras-García, A. J. Cohen, W. Yang, *J. Am. Chem. Soc.* **2010**, *132*(18), 6498–6506.

[184] A. Otero-de-la Roza, E. R. Johnson, J. Contreras-García, *Phys. Chem. Chem. Phys.* **2012**, *14*, 12165–12172.

[185] A. D. Becke, K. E. Edgecombe, *J. Chem. Phys.* **1990**, *92*(9), 5397–5403.

[186] M. Kohout, *Faraday Discuss.* **2007**, *135*, 43–54.

[187] M. Kohout, *Int. J. Quant. Chem.* **2004**, *97*(1), 651–658.

[188] G. Rubez, J.-M. Etancelin, X. Vigouroux, M. Krajecki, J.-C. Boisson, E. Héon, *J. Comput. Chem.* **2017**, *38*(14), 1071–1083.

[189] F. Weinhold, C. R. Landis, E. D. Glendening, *Int. J. Quant. Chem.* **2016**, *35*(3), 399–440.

[190] A. E. Reed, F. Weinhold, *J. Chem. Phys.* **1983**, *78*(6), 4066–4073.

[191] J. P. Foster, F. Weinhold, *J. Am. Chem. Soc.* **1980**, *102*(24), 7211–7218.

[192] G. N. Lewis, *J. Am. Chem. Soc.* **1916**, *38*(4), 762–785.

[193] I. Langmuir, *J. Am. Chem. Soc.* **1919**, *41*(10), 1543–1559.

[194] T. Ziegler, A. Rauk, *Theor. Chim. Acta* **1977**, *46*(1), 1–10.

[195] M. Von Hopffgarten, G. Frenking, *Rev.: Comput. Mol. Sci* **2012**, *2*, 43.

[196] G. C. Terstappen, C. Schlüpen, R. Raggiaschi, G. Gaviraghi, *Nat. Rev. Drug Discov.* **2007**, *6*(11), 891–903.

[197] D. di Bernardo, M. J. Thompson, T. S. Gardner, S. E. Chobot, E. L. Eastwood, A. P. Wojtovich, S. J. Elliott, S. E. Schaus, J. J. Collins, *Nat. Biotechnol.* **2005**, *23*(3), 377–383.

[198] A. Ambesi-Impiombato, D. d. Bernardo, *Curr. Bioinform.* **2006**, *1*, 3–13.

[199] Q. Liu, C. K. Kwoh, J. Li, *J. Chem. Inf. Model.* **2013**, *53*(11), 3076–3085.

[200] D. Rodbard, H. A. Feldman, in *Methods in Enzymology*, vol. 36, Academic Press, **1975**, 3–16.

[201] G. Weber, C. B. Anfinsen, J. T. Edsall, F. M. Richards, in *Advances in Protein Chemistry*, vol. 29, Academic Press, **1975**, 1–83.

[202] A. J. McCoy, V. Chandana Epa, P. M. Colman, *J. Mol. Bio.* **1997**, *268*(2), 570–584.

[203] L. Gráf, A. Jancsó, L. Szilágyi, G. Hegyi, K. Pintér, G. Náray-Szabó, J. Hepp, K. Medzihradszky, W. J. Rutter, *P. Natl. Acad. Sci. USA* **1988**, *85*, 4961–4965.

[204] H. A. Gabb, R. M. Jackson, M. J. E. Sternberg, *J. Mol. Bio.* **1997**, *272*(1), 106–120.

[205] A. Bjorklund, C. Svendsen, *Nature* **1999**, *397*(6720), 569–570.

[206] R. D. Broadwell, B. J. Balin, M. Salcman, R. S. Kaplan, *P. Natl. Acad. Sci. USA* **1983**, *80*, 7352–7356.

[207] T. H. Schneider, M. Rieger, K. Ansorg, A. N. Sobolev, T. Schirmeister, B. Engels, S. Grabowsky, *New J. Chem.* **2015**, *39*(7), 5841–5853.

[208] E. Fischer, *Ber. Dtsch. Chem. Ges.* **1894**, *27*(3), 2985–2993.

[209] D. E. Koshland, *P. Natl. Acad. Sci. USA* **1958**, *44*(2), 98–104.

[210] H. F. Fisher, N. Singh, in *Methods in Enzymology*, vol. 259, Academic Press, **1995**, 194–221.

[211] A. Wortmann, M. C. Jecklin, D. Touboul, M. Badertscher, R. Zenobi, *J. Mass Spectrom.* **2008**, *43*(5), 600–608.

[212] B. Meyer, T. Peters, *Angew. Chem. Int. Ed.* **2003**, *42*(8), 864–890.

[213] D. A. Annis, N. Nazef, C.-C. Chuang, M. P. Scott, H. M. Nash, *J. Am. Chem. Soc.* **2004**, *126*(47), 15495–15503.

[214] J. S. Soothill, R. Ward, A. J. Girling, *J. Antimicrob. Chemother.* **1992**, *29*(2), 137–139.

[215] G. W. Caldwell, Z. Yan, W. Lang, J. A. Masucci, *Curr. Top. Med. Chem.* **2012**, *12*(11), 1282–1290.

[216] W. Rut, K. Groborz, L. Zhang, X. Sun, M. Zmudzinski, R. Hilgenfeld, M. Drag, *bioRxiv* **2020**.

[217] L. Zhang, D. Lin, X. Sun, U. Curth, C. Drosten, L. Sauerhering, S. Becker, K. Rox, R. Hilgenfeld, *Science* **2020**, *368*(6489), 409–412.

[218] Y.-Y. Zheng, Y.-T. Ma, J.-Y. Zhang, X. Xie, *Nat. Rev. Cardiol.* **2020**, *17*(5), 259–260.

[219] P. Mo, Y. Xing, Y. Xiao, L. Deng, Q. Zhao, H. Wang, Y. Xiong, Z. Cheng, S. Gao, K. Liang, M. Luo, T. Chen, S. Song, Z. Ma, X. Chen, R. Zheng, Q. Cao, F. Wang, Y. Zhang, *Clin. Infect. Dis.* **2020**, ciaa270.

[220] A. Mullard, *Nat. Rev. Drug Discov.* **2015**, *14*(2), 77–81.

[221] S. A. Sundberg, *Curr. Opin. Biotechnol.* **2000**, *11*(1), 47–53.

[222] R. Potyrailo, K. Rajan, K. Stoewe, I. Takeuchi, B. Chisholm, H. Lam, *ACS Comb. Sci.* **2011**, *13*(6), 579–633.

[223] P. J. Gane, P. M. Dean, *Curr. Opin. Struct. Biol.* **2000**, *10*(4), 401–404.

[224] H. A. Carlson, J. A. McCammon, *Mol. Pharmacol.* **2000**, *57*(2), 213.

[225] S. Mandal, M. Moudgil, S. K. Mandal, *Eur. J. Pharmacol.* **2009**, *625*(1), 90–100.

[226] W. Duch, K. Swaminathan, J. Meller, *Curr. Pharm. Des.* **2007**, *13*(14), 1497–1508.

[227] R. L. Hilderbrandt, *Comput. Chem.* **1977**, *1*(3), 179–186.

[228] D. N. J. White, *Comput. Chem.* **1977**, *1*(3), 225–233.

[229] J. W. Ponder, D. A. Case, in *Advances in Protein Chemistry*, vol. 66, Academic Press, **2003**, 27–85.

[230] C. F. Wong, J. A. McCammon, in *Advances in Protein Chemistry*, vol. 66, Academic Press, **2003**, 87–121.

[231] M. González, *École thématique de la Société Française de la Neutronique* **2011**, *12*, 169–200.

[232] O. Guvench, A. D. MacKerell, in *Molecular Modeling of Proteins*, (Published by A. Kukol), Humana Press, Totowa, NJ, **2008**, 63–88.

[233] A. K. Rappe, C. J. Casewit, K. S. Colwell, W. A. Goddard, W. M. Skiff, *J. Am. Chem. Soc.* **1992**, *114*(25), 10024–10035.

[234] S. Grimme, *J. Chem. Theory Comput.* **2014**, *10*(10), 4497–4514.

[235] A. D. MacKerell, D. Bashford, M. Bellott, R. L. Dunbrack, J. D. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F. T. K. Lau, C. Mattos, S. Michnick, T. Ngo, D. T. Nguyen, B. Prodhom, W. E. Reiher, B. Roux, M. Schlenkrich, J. C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiórkiewicz-Kuczera, D. Yin, M. Karplus, *J. Phys. Chem. B* **1998**, *102*(18), 3586–3616.

[236] A. D. MacKerell, M. Feig, C. L. Brooks, *J. Am. Chem. Soc.* **2004**, *126*(3), 698–699.

[237] R. B. Best, X. Zhu, J. Shim, P. E. M. Lopes, J. Mittal, M. Feig, A. D. MacKerell, *J. Chem. Theory Comput.* **2012**, *8*(9), 3257–3273.

[238] A. Pullman, B. Pullman, M. Florkin, E. Stotz, in *Comprehensive Biochemistry*, vol. 22, Elsevier, **1967**, 1–60.

[239] C. F. Matta, L. Massa, *Future Med. Chem.* **2011**, *3*(16), 1971–1974.

[240] A. A. Arabi, C. F. Matta, *Future Med. Chem.* **2016**, *8*(4), 361–371.

[241] C. F. Matta, *Quantum Biochemistry*, vol. 1, Wiley-VCH John Wiley, distributor, Weinheim Chichester, **2010**.

[242] C. F. Matta, *Quantum Biochemistry*, vol. 2, Wiley-VCH John Wiley, distributor, Weinheim Chichester, **2010**.

[243] C. F. Matta, *J. Comput. Chem.* **2014**, *35*(16), 1165–1198.

[244] I. Sumar, P. W. Ayers, C. F. Matta, *Chem. Phys. Lett.* **2014**, *612*, 190 – 197.

[245] I. Sumar, R. Cook, P. W. Ayers, C. F. Matta, *Comput. Theor. Chem.* **2015**, *1070*, 55 – 67.

[246] I. Sumar, R. Cook, P. W. Ayers, C. F. Matta, *Phys. Scripta* **2015**, *91*(1), 013001.

[247] C. F. Matta, *Comput. Theor. Chem.* **2018**, *1124*, 1 – 14.

[248] M. Levitt, *J. Mol. Bio.* **1983**, *170*(3), 723–764.

[249] W. C. Swope, J. W. Pitera, F. Suits, *J. Phys. Chem. B* **2004**, *108*(21), 6571–6581.

[250] J. D. Chodera, W. C. Swope, J. W. Pitera, K. A. Dill, *Multiscale Model. Sim.* **2006**, *5*(4), 1214–1226.

[251] V. Salmaso, S. Moro, *Front. Pharmacol.* **2018**, *9*, 923.

[252] M. De Vivo, M. Masetti, G. Bottegoni, A. Cavalli, *J. Med. Chem.* **2016**, *59*(9), 4035–4061.

[253] V. Salmaso, M. Sturlese, A. Cuzzolin, S. Moro, *Structure* **2017**, *25*(4), 655–662.

[254] S. Tsuneyuki, H. Aoki, M. Tsukada, Y. Matsui, *Phys. Rev. Lett.* **1990**, *64*, 776–779.

[255] R. Tycko, G. Dabbagh, R. M. Fleming, R. C. Haddon, A. V. Makhija, S. M. Zahurak, *Phys. Rev. Lett.* **1991**, *67*(14), 1886–1889.

[256] Y. Shibuta, T. Suzuki, *J. Chem. Phys.* **2008**, *129*(14), 144102.

[257] J. C. Phillips, Y. Sun, N. Jain, E. J. Bohm, L. V. Kale, in *SC14: International Conference for High Performance Computing, Networking, Storage and Analysis*, IEEE, New Orleans, Louisiana, **2014**, –.

[258] M. Lu, G. Hou, H. Zhang, C. L. Suiter, J. Ahn, I.-J. L. Byeon, J. R. Perilla, C. J. Langmead, I. Hung, P. L. Gor'kov, Z. Gan, W. Brey, C. Aiken, P. Zhang, K. Schulten, A. M. Gronenborn, T. Polenova, *P. Natl. Acad. Sci. USA* **2015**, *112*(47), 14617.

[259] A. C. T. van Duin, A. Strachan, S. Stewman, Q. Zhang, X. Xu, W. A. Goddard, *J. Phys. Chem. A* **2003**, *107*(19), 3803–3811.

[260] D. W. Brenner, B. J. Garrison, *Phys. Rev. B* **1986**, *34*, 1304–1307.

[261] B. W. H. van Beest, G. J. Kramer, R. A. van Santen, *Phys. Rev. Lett.* **1990**, *64*, 1955–1958.

[262] W. G. J. Hol, *Angew. Chem. Int. Ed. Engl.* **1986**, *25*(9), 767–778.

[263] T. L. Blundell, H. Jhoti, C. Abell, *Nat. Rev. Drug Discov.* **2002**, *1*(1), 45–54.

[264] S. P. Williams, L. F. Kuyper, K. H. Pearce, *Curr. Opin. Chem. Biol.* **2005**, *9*(4), 371–380.

[265] A. M. Davis, S. J. Teague, G. J. Kleywegt, *Angew. Chem. Int. Ed.* **2003**, *42*(24), 2718–2736.

[266] J. R. Helliwell, *Rep. Prog. Phys.* **1984**, *47*(11), 1403–1497.

[267] I. Usón, G. M. Sheldrick, *Curr. Opin. Struct. Biol.* **1999**, *9*(5), 643–648.

[268] E. E. Lattman, *Proteins: Structure, Function, and Genetics* **1994**, *18*(2), 103–106.

[269] V. L. Nienaber, P. L. Richardson, V. Klighofer, J. J. Bouska, V. L. Giranda, J. Greer, *Nat. Biotechnol.* **2000**, *18*(10), 1105–1108.

[270] C. L. M. J. Verlinde, W. G. J. Hol, *Structure* **1994**, *2*(7), 577–587.

[271] P. J. Goodford, *J. Med. Chem.* **1985**, *28*(7), 849–857.

[272] K. N. Allen, C. R. Bellamacina, X. Ding, C. J. Jeffery, C. Mattos, G. A. Petsko, D. Ringe, *J. Phys. Chem.* **1996**, *100*(7), 2605–2611.

[273] D. Mendel, V. W. Cornish, P. G. Schultz, *Annu. Rev. Bioph. Biom.* **1995**, *24*(1), 435–462.

[274] G. Klebe, *J. Mol. Bio.* **1994**, *237*(2), 212–235.

[275] C. Pascard, *Acta Cryst.* **1995**, *D51*(4), 407–417.

[276] M. W. Shi, S. G. Stewart, A. N. Sobolev, B. Dittrich, T. Schirmeister, P. Luger, M. Hesse, Y.-S. Chen, P. R. Spackman, M. A. Spackman, S. Grabowsky, *J. Phys. Org. Chem.* **2017**, *30*(11), e3683.

[277] A. J. Stone, I. C. Hayes, *Faraday Discuss. Chem. Soc.* **1982**, *73*, 19–31.

[278] K. Szalewicz, K. Patkowski, B. Jeziorski, *Intermolecular Interactions via Perturbation Theory: From Diatoms to Biomolecules*, Springer, Berlin, Heidelberg, chap. 3, **2005**, 43–117.

[279] I. J. Bruno, J. C. Cole, J. P. M. Lommerse, R. S. Rowland, R. Taylor, M. L. Verdonk, *J. Comput. Aid. Mol. Des.* **1997**, *11*(6), 525–537.

[280] P. R. Spackman, L.-J. Yu, C. J. Morton, M. W. Parker, C. S. Bond, M. A. Spackman, D. Jayatilaka, S. P. Thomas, *Angew. Chem. Int. Edit.* **2019**, *58*(47), 16780–16784.

[281] D. E. Hibbs, C. J. Austin-Woods, J. A. Platts, J. Overgaard, P. Turner, *Chem.–Eur. J.* **2003**, *9*(5), 1075–1084.

[282] P. Luger, *Org. Biomol. Chem.* **2007**, *5*(16), 2529–2540.

[283] S. Grabowsky, T. Pfeuffer, L. Chęcińska, M. Weber, W. Morgenroth, P. Luger, T. Schirmeister, *Eur. J. Org. Chem.* **2007**, *2007*(17), 2759–2768.

[284] S. Grabowsky, D. Jayatilaka, R. F. Fink, T. Schirmeister, B. Engels, *Z. Anorg. Allg. Chem.* **2013**, *639*(11), 1905–1921.

[285] G. Jeffrey, *The Application of Charge Density Research to Chemistry and Drug Design*, Springer US, Boston, MA, **1991**.

[286] C. Gatti, P. Macchi, *Modern Charge-Density Analysis*, Springer-Verlag GmbH, Dordrecht, **2012**, 1–78.

[287] M. Mladenovic, M. Arnone, R. F. Fink, B. Engels, *J. Phys. Chem. B* **2009**, *113*(15), 5072–5082.

[288] M. A. Spackman, P. G. Byrom, M. Alfredsson, K. Hermansson, *Acta Cryst.* **1999**, *A55*(1), 30–47.

[289] B. Dittrich, M. A. Spackman, *Acta Cryst.* **2007**, *A63*(5), 426–436.

[290] N. Brown, *Bioisosterism in medicinal chemistry*, vol. 54, Wiley Online Library, **2012**.

[291] N. A. Meanwell, *J. Med. Chem.* **2011**, *54*(8), 2529–2591.

[292] C. F. Matta, A. A. Arabi, D. F. Weaver, *Eur. J. Med. Chem.* **2010**, *45*(5), 1868 – 1872.

[293] B. Dittrich, C. F. Matta, *IUCrJ* **2014**, *1*(6), 457–469.

[294] R. Coulombe, P. Grochulski, J. Sivaraman, R. Ménard, J. S. Mort, M. Cygler, *EMBO J.* **1996**, *15*(20), 5492–5503.

[295] H. A. Chapman, R. J. Riese, G.-P. Shi, *Annu. Rev. Physiol.* **1997**, *59*(1), 63–88.

[296] P. Pyrko, A. Kardosh, W. Wang, W. Xiong, A. H. Schönthal, T. C. Chen, *Cancer Res.* **2007**, *67*(22), 10920–10928.

[297] World Health Organization, *WHO model list of essential medicines: 20th list* **2017**.

[298] M. Baumann, I. R. Baxendale, *Reaction Chemistry & Engineering* **2016**, *1*(2), 147–150.

[299] K. D. Rainsford, *Ibuprofen*, John Wiley & Sons, Ltd, **2015**.

[300] M. T. Donnelly, C. J. Hawkey, *Aliment. Pharm. Therap.* **1997**, *11*(2), 227–235.

[301] R. A. Momin, D. L. De Witt, M. G. Nair, *Phytother. Res.* **2003**, *17*(8), 976–979.

[302] S. H. Yalkowsky, Y. He, P. Jain, *Handbook of aqueous solubility data*, CRC press, **2016**.

[303] M. Moir, J. J. Danon, T. A. Reekie, M. Kassiou, *Expert Opin. Drug Discovery* **2019**, *14*(11), 1137–1149.

[304] T. Cernak, K. D. Dykstra, S. Tyagarajan, P. Vachal, S. W. Krska, *Chem. Soc. Rev.* **2016**, *45*, 546–576.

[305] M. C. White, J. Zhao, *J. Am. Chem. Soc.* **2018**, *140*(43), 13988–14009.

[306] R. Hoffmann, *Angew. Chem. Int. Edit.* **1982**, *21*(10), 711–724.

[307] D. Seebach, *Angew. Chem. Int. Edit.* **1979**, *18*(4), 239–258.

[308] M. Geyer, E. Wellner, U. Jurva, S. Saloman, D. Armstrong, R. Tacke, *ChemMedChem* **2015**, *10*(5), 911–924.

[309] W. Bains, R. Tacke, *Curr. Opin. Drug Disc.* **2003**, *6*(4), 526–543.

[310] R. Tacke, S. Dörrich, in *Atypical Elements in Drug Design*, (Published by J. Schwarz), Springer International Publishing, Cham, **2016**, 29–59.

[311] A. K. Franz, S. O. Wilson, *J. Med. Chem.* **2013**, *56*(2), 388–405.

[312] N. F. Lazareva, I. M. Lazarev, *Russ. Chem. B.* **2015**, *64*(6), 1221–1232.

[313] N. Takashi, K. Isogai, N. Katunuma, Y. Tarumoto, M. Ohzeki, *J. Biochem.-Tokyo* **1981**, *90*(3), 893–896.

[314] T. Towatari, T. Nikawa, M. Murata, C. Yokoo, M. Tamai, K. Hanada, N. Katunuma, *FEBS Lett.* **1991**, *280*(2), 311–315.

[315] N. Schaschke, I. Assfalg-Machleidt, W. Machleidt, D. Turk, L. Moroder, *Bioorg. Med. Chem.* **1997**, *5*(9), 1789–1797.

[316] N. Schaschke, D. Deluca, I. Assfalg-Machleidt, C. Höhneke, C. P. Sommerhoff, W. Machleidt, *Biol. Chem.* **2002**, *383*(5), 849–852.

[317] B. J. Gour-Salin, P. Lachance, C. Plouffe, A. C. Storer, R. Menard, *J. Med. Chem.* **1993**, *36*(6), 720–725.

[318] B. J. Goursalin, P. Lachance, P. R. Bonneau, A. C. Storer, H. Kirschke, D. Broemme, *Bioorg. Chem.* **1994**, *22*(3), 227–241.

[319] A. J. Barrett, A. A. Kembhavi, M. A. Brown, H. Kirschke, C. G. Knight, M. Tamai, K. Hanada, *Biochem. J.* **1982**, *201*(1), 189–198.

[320] G. Dodson, A. Wlodawer, *Trends Biochem. Sci.* **1998**, *23*(9), 347–352.

[321] A. J. Beveridge, *Protein Sci.* **1996**, *5*(7), 1355–1365.

[322] W. H. Miller, *A Tract on Crystallography: Designed for the Use of Students in the University*, Deighton, Bell and Company, **1863**.

[323] W. H. Bragg, W. L. Bragg, *The Crystalline State: The optical principles of the diffraction of X-rays, by RW James*, vol. 2, G. Bell, **1948**.

[324] P. Debye, *Ann. Phys.* **1914**, *43*, 49.

[325] I. Waller, *Z. Phys.* **1923**, *17*(1), 398–408.

[326] D. Sayre, *Acta Cryst.* **1952**, *5*(1), 60–65.

[327] J. Karle, H. Hauptman, *Acta Cryst.* **1950**, *3*(3), 181–187.

[328] J. Karle, *Acta Cryst.* **1971**, *B27*(11), 2063–2065.

[329] G. M. Sheldrick, *Acta Cryst.* **2015**, *A71*(1), 3–8.

[330] A. van der Lee, *J. Appl. Crystallogr.* **2013**, *46*(5), 1306–1315.

[331] G. Oszlányi, A. Süto, *Acta Cryst.* **2008**, *A64*(1), 123–134.

[332] G. C. Ford, J. S. Rollett, *Acta Cryst.* **1970**, *A26*(1), 162.

[333] Z. Su, P. Coppens, *Acta Cryst.* **1997**, *A53*(6), 749–762.

[334] F. L. Hirshfeld, *Acta Cryst.* **1976**, *A32*(2), 239–244.

[335] R. F. Stewart, *J. Chem. Phys.* **1969**, *51*(10), 4569–4577.

[336] J. Bentley, R. F. Stewart, *J. Chem. Phys.* **1975**, *63*(9), 3794–3803.

[337] R. F. Stewart, *Acta Cryst.* **1976**, *A32*(4), 565–574.

[338] R. F. Stewart, *Isr. J. Chem.* **1977**, *16*(23), 111–114.

[339] R. F. Stewart, *Isr. J. Chem.* **1977**, *16*(23), 124–131.

[340] A. D. Becke, *J. Chem. Phys.* **1988**, *88*(4), 2547–2553.

[341] V. I. Lebedev, *USSR Computational Mathematics and Mathematical Physics* **1976**, *16*(2), 10–24.

[342] R. Ditchfield, W. J. Hehre, J. A. Pople, *J. Chem. Phys.* **1971**, *54*(2), 724–728.

[343] T. H. Dunning, *J. Chem. Phys.* **1989**, *90*(2), 1007–1023.

[344] F. Weigend, R. Ahlrichs, *Phys. Chem. Chem. Phys.* **2005**, *7*(18), 3297–3305.

[345] K. L. Schuchardt, B. T. Didier, T. Elsethagen, L. Sun, V. Gurumoorthi, J. Chase, J. Li, T. L. Windus, F. Weigend, F. Furche, R. Ahlrichs, *J. Chem. Phys.* **2003**, *119*, 12753–12762.

[346] S. Sadjadi, C. F. Matta, K. H. Lemke, I. P. Hamilton, *J. Phys. Chem. A* **2011**, *115*(45), 13024–13035.

[347] T. A. Keith, M. J. Frisch, *J. Phys. Chem. A* **2011**, *115*(45), 12879–12894.

[348] S. Sadjadi, C. F. Matta, I. Hamilton, *Can. J. Chem.* **2013**, *91*(7), 583–590.

[349] K. Burke, L. O. Wagner, *Int. J. Quant. Chem.* **2013**, *113*(2), 96–101.

[350] P. A. M. Dirac, *Math. Proc. Cambridge Philos. Soc.* **1930**, *26*(3), 376–385.

[351] T. Chachiyo, *J. Chem. Phys.* **2016**, *145*(2), 021101.

[352] M. Gell-Mann, K. A. Brueckner, *Phys. Rev.* **1957**, *106*, 364–368.

[353] D. M. Ceperley, B. J. Alder, *Phys. Rev. Lett.* **1980**, *45*, 566–569.

[354] S. H. Vosko, L. Wilk, M. Nusair, *Can. J. Phys.* **1980**, *58*(8), 1200–1211.

[355] J. P. Perdew, A. Zunger, *Phys. Rev. B* **1981**, *23*, 5048–5079.

[356] L. A. Cole, J. P. Perdew, *Phys. Rev. A* **1982**, *25*, 1265–1271.

[357] J. P. Perdew, Y. Wang, *Phys. Rev. B* **1992**, *45*, 13244–13249.

[358] G. L. Oliver, J. P. Perdew, *Phys. Rev. A* **1979**, *20*, 397–403.

[359] J. P. Perdew, K. Burke, M. Ernzerhof, *Phys. Rev. Lett.* **1996**, *77*, 3865–3868.

[360] A. D. Becke, *Phys. Rev. A* **1988**, *38*, 3098–3100.

[361] C. Lee, W. Yang, R. G. Parr, *Phys. Rev. B* **1988**, *37*, 785–789.

[362] A. D. Becke, *J. Chem. Phys.* **1993**, *98*(7), 5648–5652.

[363] J. P. Perdew, M. Ernzerhof, K. Burke, *J. Chem. Phys.* **1996**, *105*(22), 9982–9985.

[364] A. D. Becke, *J. Chem. Phys.* **1993**, *98*(2), 1372–1377.

[365] C. Adamo, V. Barone, *J. Chem. Phys.* **1999**, *110*(13), 6158–6170.
[366] A. J. Cohen, P. Mori-Sánchez, W. Yang, *Chem. Rev.* **2012**, *112*(1), 289–320.
[367] E. G. Hohenstein, S. T. Chill, C. D. Sherrill, *J. Chem. Theory Comput.* **2008**, *4*(12), 1996–2000.
[368] Y. Zhao, D. G. Truhlar, *Theor. Chem. Acc.* **2008**, *120*(1), 215–241.
[369] J. Tao, J. P. Perdew, V. N. Staroverov, G. E. Scuseria, *Phys. Rev. Lett.* **2003**, *91*, 146401.
[370] V. N. Staroverov, G. E. Scuseria, J. Tao, J. P. Perdew, *J. Chem. Phys.* **2003**, *119*(23), 12129–12137.
[371] S. Grimme, *J. Chem. Phys.* **2006**, *124*(3), 034108.
[372] M. G. Medvedev, I. S. Bushmarinov, J. Sun, J. P. Perdew, K. A. Lyssenko, *Science* **2017**, *355*(6320), 49–52.
[373] K. P. Kepp, *Science* **2017**, *356*(6337), 496–496.
[374] M. G. Medvedev, I. S. Bushmarinov, J. Sun, J. P. Perdew, K. A. Lyssenko, *Science* **2017**, *356*(6337), 496–496.
[375] M. Challacombe, E. Schwegler, J. Almlöf, *J. Chem. Phys.* **1996**, *104*(12), 4685–4698.
[376] K. R. Glaesemann, M. S. Gordon, *J. Chem. Phys.* **1998**, *108*(24), 9959–9969.
[377] R. A. Kendall, H. A. Früchtl, *Theor. Chem. Acc.* **1997**, *97*(1), 158–163.
[378] E. J. Baerends, D. E. Ellis, P. Ros, *Chem. Phys.* **1973**, *2*(1), 41–51.
[379] O. Vahtras, J. Almlöf, M. W. Feyereisen, *Chem. Phys. Lett.* **1993**, *213*(5-6), 514–518.
[380] F. Weigend, M. Kattannek, R. Ahlrichs, *J. Chem. Phys.* **2009**, *130*(16), 164106.
[381] F. Neese, F. Wennmohs, A. Hansen, U. Becker, *Chem. Phys.* **2009**, *356*(1), 98–109.
[382] S. Kossmann, F. Neese, *Chem. Phys. Lett.* **2009**, *481*(4-6), 240–243.
[383] N. Castillo, C. F. Matta, R. J. Boyd, *Chem. Phys. Lett.* **2005**, *409*(4), 265 – 269.
[384] M. P. C. M. Krijn, H. Graafsma, D. Feil, *Acta Cryst.* **1988**, *B44*(6), 609–616.
[385] T. A. Jones, J.-Y. Zou, S. W. Cowan, M. Kjeldgaard, *Acta Cryst.* **1991**, *A47*(2), 110–119.
[386] J. Contreras-García, R. A. Boto, F. Izquierdo-Ruiz, I. Reva, T. Woller, M. Alonso, *Theor. Chem. Acc.* **2016**, *135*(10), 242.
[387] P. Wu, R. Chaudret, X. Hu, W. Yang, *J. Chem. Theory Comput.* **2013**, *9*(5), 2226–2234.
[388] H. Weinstein, *Int. J. Quant. Chem.* **1975**, *9*(S2), 59–69.
[389] F. Fuster, B. Silvi, *Theor. Chem. Acc.* **2000**, *104*(1), 13–21, cited By 132.
[390] F. R. Wagner, M. Kohout, Y. Grin, *J. Phys. Chem. A* **2008**, *112*(40), 9814–9828.
[391] S. Mebs, M. A. Chilleck, *Chem. Phys. Lett.* **2014**, *591*, 1–4.
[392] M. Fugel, Dissertation, University of Bremen, **2019**.
[393] S. Raub, G. Jansen, *Theor. Chem. Acc.* **2001**, *106*(3), 223–232.
[394] A. Gavezzotti, *Z. Krist.-Cryst. Mater.* **01 May. 2005**, *220*(5-6), 499–510.
[395] A. D. Bond, *J. Appl. Crystallogr.* **2014**, *47*(5), 1777–1780.
[396] M. J. Turner, S. Grabowsky, D. Jayatilaka, M. A. Spackman, *J. Phys. Chem. Lett.* **2014**, *5*(24), 4249–4255.
[397] P. Su, H. Li, *J. Chem. Phys.* **2009**, *131*(1), 014102.
[398] A. J. Thakkar, C. Lupinetti, *Atomic Polarizabilities and Hyperpolarizabilities: A critical Compilation*, published by Imperial College Press and disctributed by Scientific Publishing Co., vol. 1 von *Computational, Numerical and Mathematical Methods in Sciences and Engineering*, chap. , **2006**, 505–529.
[399] S. Grimme, J. Antony, S. Ehrlich, H. Krieg, *J. Chem. Phys.* **2010**, *132*(15), 154104.
[400] S. Grimme, S. Ehrlich, L. Goerigk, *J. Comput. Chem.* **2011**, *32*(7), 1456–1465.
[401] F. M. Bickelhaupt, E. J. Baerends, *Kohn-Sham Density Functional Theory: Predicting and Understanding Chemistry*, John Wiley & Sons, Ltd, New York, **2007**, 1–86.
[402] A. E. Reed, R. B. Weinstock, F. Weinhold, *J. Chem. Phys.* **1985**, *83*(2), 735–746.
[403] E. D. Glendening, F. Weinhold, *J. Comput. Chem.* **1998**, *19*(6), 593–609.
[404] E. D. Glendening, F. Weinhold, *J. Comput. Chem.* **1998**, *19*(6), 610–627.
[405] E. D. Glendening, J. K. Badenhoop, F. Weinhold, *J. Comput. Chem.* **1998**, *19*(6), 628–646.
[406] W. C. Price, S. S. Chissick, T. Ravensdale, Wave mechanics: the first fifty years, **1973**.
[407] K. R. Roby, *Mol. Phys.* **1974**, *27*(1), 81–104.
[408] K. R. Roby, *Mol. Phys.* **1974**, *28*(6), 1441–1456.
[409] K. R. Roby, *Theor. Chim. Acta* **1974**, *33*(2), 105–113.
[410] M. D. Gould, C. Taylor, S. K. Wolff, G. S. Chandler, D. Jayatilaka, *Theor. Chem. Acc.* **2008**, *119*(1), 275–290.

[411] T. Darden, D. York, L. Pedersen, *J. Chem. Phys.* **1993**, *98*(12), 10089–10092.

[412] P. P. Ewald, *Ann. Phys.* **1921**, *369*(3), 253–287.

[413] L. Verlet, *Phys. Rev.* **1967**, *159*, 98–103.

[414] J.-P. Ryckaert, G. Ciccotti, H. J. C. Berendsen, *J. Comput. Chem.* **1977**, *23*(3), 327–341.

[415] W. C. Swope, H. C. Andersen, P. H. Berens, K. R. Wilson, *J. Chem. Phys.* **1982**, *76*(1), 637–649.

[416] D. Porezag, T. Frauenheim, T. Köhler, G. Seifert, R. Kaschner, *Phys. Rev. B* **1995**, *51*, 12947–12957.

[417] W. F. van Gunsteren, P. K. Weiner, A. J. Wilkinson, *Computer simulation of biomolecular systems: theoretical and experimental applications*, vol. 3, Springer Science & Business Media, **2013**.

[418] S. B. Dixit, C. Chipot, *J. Phys. Chem. A* **2001**, *105*(42), 9795–9799.

[419] D. L. Beveridge, F. M. DiCapua, *Annu. Rev. Biophys. Bio.* **1989**, *18*(1), 431–492.

[420] W. F. van Gunsteren, X. Daura, A. E. Mark, *Helv. Chim. Acta* **2002**, *85*(10), 3113–3129.

[421] J. Gao, K. Kuczera, B. Tidor, M. Karplus, *Science* **1989**, *244*(4908), 1069–1072.

[422] A. Warshel, M. Levitt, *J. Mol. Bio.* **1976**, *103*(2), 227–249.

[423] M. J. Turner, J. J. McKinnon, S. K. Wolff, D. J. Grimwood, P. R. Spackman, D. Jayatilaka, M. A. Spackman, CrystalExplorer17, University of Western Australia, **2017**.

[424] M. Fugel, D. Jayatilaka, E. Hupf, J. Overgaard, V. R. Hathwar, P. Macchi, M. J. Turner, J. A. K. Howard, O. V. Dolomanov, H. Puschmann, B. B. Iversen, H.-B. Bürgi, S. Grabowsky, *IUCrJ* **2018**, *5*(1), 32–44.

[425] L. A. Malaspina, A. Genoni, S. Grabowsky, *J. Appl. Cryst.* **2020**.

[426] L. Midgley, L. J. Bourhis, O. Dolomanov, N. Peyerimhoff, H. Puschmann, *arXiv: 1911.08847v1* **2019**.

[427] T. Koga, K. Kanayama, T. Watanabe, T. Imai, A. J. Thakkar, *Theor. Chem. Acc.* **2000**, *104*(5), 411–413.

[428] T. Koga, K. Kanayama, S. Watanabe, A. J. Thakkar, *Int. J. Quant. Chem.* **1999**, *71*(6), 491–497.

[429] R. Bast, Numgrid: Numerical integration grid for molecules, **2020**.

[430] J. Burkardt, SPHERE LEBEDEV RULE: Quadrature Rules for the Unit Sphere, **2010**.

[431] B. A. Hess, *Phys. Rev. A* **1986**, *33*(6), 3742.

[432] G. Jansen, B. A. Hess, *Phys. Rev. A* **1989**, *39*(11), 6016.

[433] A. Wolf, M. Reiher, B. A. Hess, *J. Chem. Phys.* **2002**, *117*(20), 9215–9226.

[434] R. Destro, R. E. Marsh, R. Bianchi, *J. Phys. Chem.* **1988**, *92*(4), 966–973.

[435] R. W. Grosse-Kunstleve, N. K. Sauter, N. W. Moriarty, P. D. Adams, *J. Appl. Crystallogr.* **2002**, *35*(1), 126–136.

[436] S. Thomas, P. R. Spackman, D. Jayatilaka, M. A. Spackman, *J. Chem. Theory Comput.* **2018**, *14*(3), 1614–1623.

[437] C. F. Mackenzie, P. R. Spackman, D. Jayatilaka, M. A. Spackman, *IUCrJ* **2017**, *4*(5), 575–587.

[438] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. A. Montgomery, J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, P. G. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J. M. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels, O. Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski, D. J. Fox, Gaussian09 Rev. D 1, Gaussian Inc., Wallingford, **2014**.

[439] T. Lu, F. Chen, *J. Comput. Chem.* **2012**, *33*(5), 580–592.

[440] J. Contreras-García, E. R. Johnson, S. Keinan, R. Chaudret, J.-P. Piquemal, D. N. Beratan, W. Yang, *J. Chem. Theory Comput.* **2011**, *7*(3), 625–632.

[441] M. Kohout, DGrid (Version 5.0), **2019**.

[442] J. Nickolls, I. Buck, M. Garland, K. Skadron, *Queue* **2008**, *6*(2), 40–53.

[443] T. A. Keith, AIMAll (Version 19.10.12), **2019**.

[444] V. March, Y. M. Teo, X. Wang, in *Proceedings of the Fifth Australasian Symposium on ACSW Frontiers*, vol. 68 von *ACSW*, Australian Computer Society, Inc., Darlinghurst, Australia, Australia, vol. 68 von *ACSW*, **2007**, 41–48.

[445] E. D. Glendening, J. K. Badenhoop, A. E. Reed, J. E. Carpenter, J. A. Bohmann, C. M. Morales, P. Karafiloglou, C. R. Landis, F. Weinhold, *Theoretical Chemistry Institute and Department of Chemistry, University of Wisconsin, Madison* **2018**.

[446] G. Henkelman, A. Arnaldsson, H. Jónsson, *Comp. Mat. Sci.* **2006**, *36*(3), 354–360.

[447] J. E. Stone, D. Gohara, G. Shi, *Comput. Sci. Eng.* **2010**, *12*(3), 66–73.

[448] C. G. J. Jacobi, *J. reine Angew. Math.* **1846**, 51–94.

[449] E. D. Glendening, C. R. Landis, F. Weinhold, *J. Am. Chem. Soc.* **2019**, *141*(10), 4156–4166.

[450] D. M. York, M. Karplus, *J. Phys. Chem. A* **1999**, *103*(50), 11060–11079.

[451] K. Takegahara, H. Harima, A. Yanase, *J. Phys. Soc. Jpn.* **2001**, *70*(5), 1190–1193.

[452] C. Abadía, P. A. Algarabel, B. García-Landa, M. R. Ibarra, A. del Moral, N. V. Kudrevatykh, P. E. Markin, *J. Phys.-Condens. Mat.* **1998**, *10*(2), 349–361.

[453] M. Hudák, D. Jayatilaka, L. Perašínová, S. Biskupič, J. Kožíšek, L. Bučinský, *Acta Cryst.* **2010**, *A66*(1), 78–92.

[454] L. Buinský, D. Jayatilaka, S. Grabowsky, *J. Phys. Chem. A* **2016**, *120*(33), 6650–6669.

[455] A. Genoni, L. H. R. Dos Santos, B. Meyer, P. Macchi, *IUCrJ* **2017**, *4*(2), 136–146.

[456] L. Bučinský, D. Jayatilaka, S. Grabowsky, *Acta Cryst.* **2019**, *A75*(5), 705–717.

[457] M. Ernst, A. Genoni, P. Macchi, *J. Mol. Struct.* **2020**, *1209*, 127975.

[458] R. Dovesi, R. Orlando, A. Erba, C. M. Zicovich-Wilson, B. Civalleri, S. Casassa, L. Maschio, M. Ferrabone, M. De La Pierre, P. D'Arco, Y. Noël, M. Causà, M. Rérat, B. Kirtman, *Int. J. Quant. Chem.* **2014**, *114*(19), 1287–1317.

[459] H. Birkedal, D. Madsen, R. H. Mathiesen, K. Knudsen, H.-P. Weber, P. Pattison, D. Schwarzenbach, *Acta Cryst.* **2004**, *A60*(5), 371–381.

[460] A. Dittmer, R. Izsák, F. Neese, D. Maganas, *Inorg. Chem.* **2019**, *58*(14), 9303–9315.

[461] B. Dittrich, E. Sze, J. J. Holstein, C. B. Hubschle, D. Jayatilaka, *Acta Cryst.* **2012**, *A68*(4), 435–442.

[462] B. Gruza, M. L. Chodkiewicz, J. Krzeszczakowska, P. M. Dominiak, *Acta Cryst.* **2020**, *A76*(1), 92–109.

[463] S. Lerdkanchanaporn, D. Dollimore, *J. Therm. Anal. Calorim.* **2005**, *49*(2), 879–886.

[464] A. J. Romero, C. T. Rhodes, *J. Pharm. Pharmacol.* **1993**, *45*(4), 258–262.

[465] A. Romero, L. Savastano, C. Rhodes, *Int. J. Pharm.* **1993**, *99*(2), 125 – 134.

[466] F. Xu, L.-X. Sun, Z.-C. Tan, J.-G. Liang, R.-L. Li, *Thermochim. Acta* **2004**, *412*(1), 33 – 57.

[467] P. Politzer, J. S. Murray, Z. Peralta-Inga, *Int. J. Quant. Chem.* **2001**, *85*(6), 676–684.

[468] V. Zoete, M. A. Cuendet, A. Grosdidier, O. Michielin, *J. Comput. Chem.* **2011**, *32*(11), 2359–2368.

[469] C. G. Mayne, J. Saam, K. Schulten, E. Tajkhorshid, J. C. Gumbart, *J. Comput. Chem.* **2013**, *34*(32), 2757–2770.

[470] B. J. Orlando, M. J. Lucido, M. G. Malkowski, *J. Struct. Bio.* **2015**, *189*(1), 62–66.

[471] L. Kalé, R. Skeel, M. Bhandarkar, R. Brunner, A. Gursoy, N. Krawetz, J. Phillips, A. Shinozaki, K. Varadarajan, K. Schulten, *J. Comput. Phys.* **1999**, *151*(1), 283–312.

[472] J. C. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R. D. Skeel, L. Kale, K. Schulten, *J. Comput. Chem.* **2005**, *26*(16), 1781–1802.

[473] B. S. Selinsky, K. Gupta, C. T. Sharkey, P. J. Loll, *Biochemistry-Moscow* **2001**, *40*(17), 5172–5180.

[474] C. Yokoyama, T. Tanabe, *Biochem. Bioph. Res. Co.* **1989**, *165*, 888–94.

[475] D. Picot, P. J. Loll, R. M. Garavito, *Nature* **1994**, *367*, 243–9.

[476] M. Hamberg, B. Samuelsson, *P. Natl. Acad. Sci. USA* **1973**, *70*, 899–903.

[477] T. Hla, K. Neilson, *P. Natl. Acad. Sci. USA* **1992**, *89*, 7384–8.

[478] Y. Mandour, H. Handoussa, N. Swilam, R. Hanafi, L. Mahran, *Int. J. Food Prop.* **2016**, *19*(10), 2377–2384.

[479] D. Riendeau, S. Charleson, W. Cromlish, J. A. Mancini, E. Wong, J. Guay, *Can. J. Physiol. Pharm.* **1997**, *75*(9), 1088–1095.

[480] D. Riendeau, M. D. Percival, C. Brideau, S. Charleson, D. Dubé, D. Ethier, J.-P. Falgueyret, R. W. Friesen, R. Gordon, G. Greig, J. Guay, J. Mancini, M. Ouellet, E. Wong, L. Xu, S. Boyce, D. Visco, Y. Girard, P. Prasit, R. Zamboni, I. W. Rodger, M. Gresser, A. W. Ford-Hutchinson, R. N. Young, C.-C. Chan, *J. Pharmacol. Exp. Ther.* **2001**, *296*(2), 558.

[481] M. Kato, S. Nishida, H. Kitasato, N. Sakata, S. Kawai, *J. Pharm. Pharmacol.* **2001**, *53*(12), 1679–1685.

[482] J. V. McCullagh, *J. Chem. Educ.* **2008**, *85*(7), 941.

[483] A. Lemmerer, S. A. Bourne, M. R. Caira, J. Cotton, U. Hendricks, L. C. Peinke, L. Trollope, *CrystEngComm* **2010**, *12*(11), 3634–3641.

[484] P. Molnár, P. Bombicz, C. Varga, L. Bereczki, E. Székely, G. Pokol, E. Fogassy, B. Simándi, *Chirality* **2009**, *21*(6), 628–636.

[485] S. Sasaki, *National Lab for High Energy Physics* **1989**.

[486] F. H. Allen, D. G. Watson, L. Brammer, A. G. Orpen, R. Taylor, *Typical interatomic distances: organic compounds*, American Cancer Society, chap. 9.5, **2006**, 790–811.

[487] H. D. Flack, *Acta Cryst.* **1983**, *A39*(6), 876–881.

[488] W. Smith, T. Forester, *J. Mol. Graph.* **1996**, *14*, 136–141.

[489] J. Tomasi, B. Mennucci, R. Cammi, *Chem. Rev.* **2005**, *105*(8), 2999–3094.

[490] R. B. Meyer, *Appl. Phys. Lett.* **1968**, *12*(9), 281–282.

[491] D. H. Dunlap, V. M. Kenkre, *Phys. Rev. B* **1986**, *34*(6), 3625.

[492] H. Fu, R. E. Cohen, *Nature* **2000**, *403*(6767), 281–283.

[493] G. Klebe, *J. Mol. Struct.-Theochem* **1994**, *308*, 53 – 89.

[494] H. Wolf, M. R. V. Jørgensen, Y.-S. Chen, R. Herbst-Irmer, D. Stalke, *Acta Cryst.* **2015**, *B71*(1), 10–19.

[495] S. Bähr, S. Brinkmann-Chen, M. Garcia-Borràs, J. M. Roberts, D. E. Katsoulis, K. N. Houk, F. H. Arnold, *Angew. Chem. Int. Ed.* **2020**, *59*(36), 15507–15511.

Declaration of Consent

on basis of Article 18 of the PromR Phil.-nat. 19

Last name, first name:	**Kleemiss, Florian**
Matriculation number:	**19-118-538**
Study program:	**Chemistry - Dissertation**
Title of the thesis:	**Development of Quantum-Crystallographic Methods for Chemical and Biochemical Applications**
Supervisor:	**PD Dr. Simon Grabowsky**

I declare herewith that this thesis is my own work and that I have not used any sources other than those stated. I have indicated the adoption of quotations as well as thoughts taken from other authors as such in the thesis. I am aware that the Senate pursuant to Article 36 paragraph 1 litera r of the University Act of 5 September 1996 and Article 69 of the University Statute of 7 June 2011 is authorized to revoke the title awarded on the basis of this thesis.

For the purpose of evaluation and verification of compliance with the declaration of originality and the regulations governing plagiarism, I hereby grant the University of Bern the right to process my personal data and to perform the acts of use this requires, in particular, to reproduce the written thesis and to store it permanently in a database, and to use said database, or to make said database available, to enable comparison with future theses submitted by others.

Bern, November 24, 2020

Florian Kleemiss

Florian Kleemiss

Curriculum Vitae

Personal

Address	Department of Chemistry and Biochemistry Freiestrasse 3, Bern 3012, Switzerland
Phone	+49 (0)1514 01 31 973
Email	florian.kleemiss@arcor.de
Date of birth	28.01.1992
Place of birth	Bremen, Germany
Nationality	German

Education

2009 – 2011	**Abitur**, Gymnasium Horn, Bremen, Germany Major subjects: Maths & Chemistry
2011 – 2014	**B. Sc. in Chemistry**, University of Bremen Title of Bachelor thesis: "*Molekulardynamik-Simulation der Grenzfläche zwischen flüssiger und fester Phase von verschiedenen Modellen für Argon und Wasser.*"
2014 – 2016	**M. Sc. in Chemistry**, University of Bremen Title of Master thesis: "*Crystal and enzyme environmental effects on the electron density of a cysteine protease inhibitor model compound.*"
2016 – 2019	**Scholarship** of the *Studienstiftung des Deutschen Volkes* for the PhD project "*Sila-Ibuprofen*"
2019 – 2020	**Doctorate**, University of Bern Title of thesis: "*Development of Quantum-Crystallographic Methods for Chemical and Biochemical Applications.*"

Employment

2012 – 2019	Tanzschule Renz und Partner, Bornstraße 64, 28159 Bremen, Germany
2016 – 2018	**Research associate** at the AG Grabowsky, University of Bremen, Germany
since 01/2018	Freelancer for **conceptualisation** and **construction** of high performance computing clusters (**HPC**)
since 06/2019	**Software developer** at Olexsys U.G., Regensburg, Germany
since 08/2019	**PhD Student** at the University of Bern, Switzerland

Skills and Languages

German	Mother tongue
English	Fluent
Spanish	Basic

C++	Experienced
Quantum Crystallography	Experienced
Quantum Chemical Calculations	Experienced
CUDA	Advanced
Python	Advanced
LaTeX	Advanced
System Administration	Advanced
Linux	Advanced
QM/MM	Advanced
MD	Advanced
HPC Construction and Maintenance	Intermediate
MS Office	Intermediate
Html	Intermediate
Fortran	Beginner

Experience related to teaching

2012 – 2016	Tutor for maths in the chemistry and biology studies, *University of Bremen*
2016 – 2017	Tutor for general chemistry for chemists and biologists, *University of Bremen*
2017 – 2019	Assistant in the quantum chemistry course, *University of Bremen*
2019	Tutor for the general chemistry lab, *University of Bern*

Conferences and meetings

06/2016	European Charge Density Meeting, *Warsaw*, *Poland*
08/2016	Robert F. Stewart school on electron density and related properties, *Nancy*, *France*
11/2016	Annual Meeting of the Crystallographic Society of Japan, *Mito*, *Japan*
11/2016	BL02B1 Workshop, *University of Tsukuba*, *Japan*
08/2017	IUCr Meeting, *Hyderabad*, *India*, "**Crystal and enyzme environmental effects on the electron density of ibuprofen and sila-ibuprofen**"
06/2018	School on Quantum Crystallography, *Erice*, *Italy*, "**Quantum crystallography of sila-ibuprofen**"
09/2018	2nd European Symposium on Chemical Bonds, *Oviedo*, *Spain*, "**X-ray constrained bonding analysis**"
07/2019	Tools for Chemical Bonding, *Bremen*, *Germany*, **Oragnizer**
02/2020	Joint Meeting of the Polish and German Crystallographic Associations, *Wroclaw*, *Poland*, "**Chemical bonding analysis based on routine X-ray diffraction experiments**"

Travel Abroad

01/2016	Research internship, *University of Western Australia*
– 03/2016	Title: "**Reaction Mechanism of the Meyer-Schuhster-Rearrangement**"
06/2016	European Charge Density Meeting, *Warsaw*, *Poland*
– 07/2016	Title: "**Crystal and enzyme environmental effects on the electron density of a cysteine protease inhibitor**"
11/2016	Synchrotron Measurement at **SPring-8**, *Hyogo*, *Japan*
07/2017	Synchrotron Measurement at **SPring-8**, *Hyogo*, *Japan*
04/2019	Synchrotron Measurement at **APS**, *Lemont*, *USA*

Awards and scholarships

- **Abitur-Award** of the German Chemical Society (GDCh)
- **Abitur-Award** of the German Physical Society (DPG)
- **Scholarship** of the **Studienstiftung des Deutschen Volkes** (Bachelor and Master)
- **Foreign scholarship** for a two-months research internship in Australia
- **Congress Scholarship** for giving a talk at the European Charge Density Meeting
- **Bridge Scholarship** for the PhD project from the University of Bremen
- **Scholarship** of the **Studienstiftung des Deutschen Volkes** for the PhD project
- **Scholarship** of the **GDK** for participation of the Erice School on quantum crystallography

Articles published in peer-reviewed journals

1. Florian Kleemiss, Aileen Justies, Daniel Duvinage, Patrick Watermann, Eric Ehrke, Kunihisa Sugimoto, Malte Fugel, Lorraine A. Malaspina, Anneke Dittmer, Torsten Kleemiss, Pim Puylaert, Nelly R. King, Anne Staubitz, Thomas M. Tzschentke, Ralf Dringen, Simon Grabowsky, Jens Beckmann, Sila-Ibuprofen, *J. Med. Chem.*, **2020**, *63*, 21, 12614-12622.
2. Florian Kleemiss, Erna K. Wieduwilt, Emanuel Hupf, Ming W. Shi, Scott G. Stewart, Dylan Jayatilaka, Michael J. Turner, Kunihisa Sugimoto, Eiji Nishibori, Tanja Schirmeister, Thomas C. Schmidt, Bernd Engels, Simon Grabowsky, Similarities and Differences between Crystal and Enzyme Environmental Effects on the Electron Density of Drug Molecules, *Chem. Eur. J.*, **2020**, DOI: 10.1002/chem.202003978.
3. Florian Kleemiss, Horst Puschmann, Oleg V. Dolomanov, Michael Bodensteiner, Norbert Peyerimhoff, Laura Midgley, Luc Bourhis, Alessandro Genoni, Lorraine A. Malaspina, Dylan Jayatilaka, Fraser White, Bernhard Grundkötter, Simon Steinhauer, Dieter Lentz, Simon Grabowsky, Accurate Crystal Structures and Chemical Properties from NoSpherA2, *Chem. Sci.*, **2020**, DOI: 10.1039/D0SC05526C.
4. Marian Olaru, Elena Rychagova, Sergey Ketkov, Yevhen Shynkarenko, Sergii Yakunin, Maksym V. Kovalenko, Artem Yablonskiy, Boris Andreev, Florian Kleemiss, Jens Beckmann, Matthias Vogt, A Small Cationic Organo-Copper Cluster as Thermally Robust Highly Photo- and Electro Luminescent Material, *J. Am. Chem. Soc.*, **2020**, *142*, 1, 373 – 381.
5. Dirk Schlüter, Florian Kleemiss, Malte Fugel, Enno Lork, Kunihisa Sugimoto, Simon Grabowsky, Jeffrey R. Harmer, Matthias Vogt, Non-Oxido-Vanadium(IV) Metalloradical Complexes with Bidentate 1,2 - Dithienylethene Ligands: Observation of Reversible Cyclization of the Ligand Scaffold in Solution, *Chem. Eur. J.*, **2020**, *26*, 6, 1335-1343.
6. Hikaru Yanai, Takumi Suzuki, Florian Kleemiss, Haruhiko Fukaya, Yasuo Dobashi, Lorraine A. Malaspina, Simon Grabowsky, Takashi Matsumoto, Chemical Bonding in Polarised Push-Pull Ethylenes, *Angew. Chem. Int. Ed.*, **2019**, *58*, 8839-8844.
7. Emanuel Hupf, Lorraine A. Malaspina, Sebastian Holsten, Florian Kleemiss, Alison J. Edwards, Jason R. Price, Valeri Kozich, Karsten Heyne, Stefan Mebs, Simon Grabowsky, Jens Beckmann, Proximity Enforced Agostic Interactions Involving Closed-Shell Coinage Metal Ions, *Inorg. Chem*, **2019**, *58*, 24, 16372 – 16378.
8. Lorraine A. Malaspina, Erna K. Wieduwilt, Justin Bergmann, Florian Kleemiss, Benjamin Meyer, Manuel F. Ruiz-López, Rumpa Pal, Emanuel Hupf, Jens Beckmann, Ross O. Piltz, Alison J. Edwards, Simon Grabowsky, Alessandro Genoni, Fast and Accurate Quantum Crystallography: From Small to Large, from Light to Heavy, *J. Phys. Chem. Lett.*, **2019**, *20*, 10, 6973-6982.
9. Malte Fugel, Maksym V. Ponomarenko, Maxie F. Hesse, Lorraine A. Malaspina, Florian Kleemiss, Kunihisa Sugimoto, Alessandro Genoni, Gerd-Volker Röschenthaler, Simon Grabowsky, Complementary bonding analysis of the N-Si interaction in pentacoordinated silicon compounds using quantum crystallography, *Dalton. Trans.*, **2019**, *48*, 16330-16339.

10. Malte Fugel, Florian Kleemiss, Lorraine A. Malaspina, Rumpa Pal, Peter R. Spackman, Dylan Jayatilaka, Simon Grabowsky, Investigating the Resonance in Nitric Acid and the Nitrate Anion Based on a Modern Bonding Analysis, *Aust. J. Chem.*, **2018**, *71*, 4, 227-237.

11. Markus Rohdenburg, Martin Mayer, Max Grellmann, Carsten Jenne, Tobias Borrmann, Florian Kleemiss, Vladimir A. Azov, Knut R. Asmis, Simon Grabowsky, Jonas Warneke Superelectrophilic Behavior of an Anion Demonstrated by the Spontaneous Binding of Noble Gases to $[B_{12}Cl_{11}]^-$, *Angew. Chem. Int. Ed.*, **2017**, *56*, 7980-7985.

12. Nikolay A. Pushkarevsky, Pavel A. Petrov, Denis S. Grigoriev, Anton I. Smolentsev, Lucia M. Lee, Florian Kleemiss, Georgy E. Salnikov, Sergey N. Konchenko, Ignacio Vargas-Baca, Simon Grabowsky, Jens Beckmann, Andrey V. Zibarev, The Nature of Bonding in Donor-Acceptor Interactions Exemplified by Complexes of N-Heterocyclic Carbenes with 1,2,5-Telluradiazoles, *Chem. Eur. J.*, **2017**, *23*, 46, 10987-10991.